

日 本 国 特 許 庁  
PATENT OFFICE  
JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日

Date of Application:

1999年11月30日

出 願 番 号

Application Number:

平成11年特許願第340472号

出 願 人

Applicant (s):

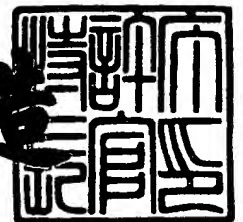
ソニー株式会社

CERTIFIED COPY OF  
PRIORITY DOCUMENT

2000年 9月 8日

特許庁長官  
Commissioner,  
Patent Office

及 川 耕 造



【書類名】 特許願

【整理番号】 9900885303

【提出日】 平成11年11月30日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 3/00  
G05B 19/4155

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社  
内

【氏名】 浅野 康治

【発明者】

【住所又は居所】 東京都品川区北品川 6 丁目 7 番 3 5 号 ソニー株式会社  
内

【氏名】 包 洪長

【特許出願人】

【識別番号】 000002185

【氏名又は名称】 ソニー株式会社

【代表者】 出井 伸之

【代理人】

【識別番号】 100082131

【弁理士】

【氏名又は名称】 稲本 義雄

【電話番号】 03-3369-6479

【手数料の表示】

【予納台帳番号】 032089

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9708842

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声処理装置および音声処理方法、並びに記録媒体

【特許請求の範囲】

【請求項 1】 ロボットに内蔵される音声処理装置であって、  
音声処理する音声処理手段と、  
前記ロボットの状態に基づいて、前記音声処理手段による音声処理を制御する  
制御手段と

を備えることを特徴とする音声処理装置。

【請求項 2】 制御手段は、前記ロボットの行動、感情、または本能の状態  
に基づいて、前記音声処理を制御する

ことを特徴とする請求項 1 に記載の音声処理装置。

【請求項 3】 前記音声処理手段は、音声合成処理を行い、合成音を出力す  
る音声合成手段で構成され、

前記制御手段は、前記ロボットの状態に基づいて、前記音声合成手段による音  
声合成処理を制御する

ことを特徴とする請求項 1 に記載の音声処理装置。

【請求項 4】 前記制御手段は、前記音声合成手段が出力する合成音の音韻  
情報または韻律情報を制御する

ことを特徴とする請求項 3 に記載の音声処理装置。

【請求項 5】 前記制御手段は、前記音声合成手段が出力する合成音の発話  
速度または音量を制御する

ことを特徴とする請求項 3 に記載の音声処理装置。

【請求項 6】 前記音声処理手段は、入力された音声の韻律情報若しくは音  
韻情報を抽出し、

前記ロボットの感情の状態が、前記韻律情報若しくは音韻情報に基づいて変更  
され、または前記ロボットが、前記韻律情報若しくは音韻情報に対応する行動を  
とる

ことを特徴とする請求項 1 に記載の音声処理装置。

【請求項 7】 前記音声処理手段は、入力された音声を認識する音声認識手

段で構成され、

前記ロボットが、前記音声認識手段が出力する音声認識結果の信頼性に対応する行動をとり、または前記ロボットの感情の状態が、前記信頼性に基づいて変更される

ことを特徴とする請求項 1 に記載の音声処理装置。

【請求項 8】 前記制御手段は、前記ロボットが行っている行動を認識し、その行動に対する負荷に基づいて、前記音声処理手段による音声処理を制御することを特徴とする請求項 1 に記載の音声処理装置。

【請求項 9】 前記ロボットは、前記音声処理手段による音声処理に割り当て可能なリソースに対応する行動をとる

ことを特徴とする請求項 8 に記載の音声処理装置。

【請求項 10】 ロボットに内蔵される音声処理装置の音声処理方法であって、

音声进行处理する音声処理ステップと、

前記ロボットの状態に基づいて、前記音声処理ステップにおける音声処理を制御する制御ステップと

を備えることを特徴とする音声処理方法。

【請求項 11】 ロボットに音声処理を行わせるために、コンピュータが実行するプログラムが記録されている記録媒体であって、

音声进行处理する音声処理ステップと、

前記ロボットの状態に基づいて、前記音声処理ステップにおける音声処理を制御する制御ステップと

を備えるプログラムが記録されている

ことを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声処理装置および音声処理方法、並びに記録媒体に関し、特に、音声認識や音声合成等の音声処理機能を有するロボットに用いて好適な音声処理

装置および音声処理方法、並びに記録媒体に関する。

【0002】

【従来の技術】

従来より、玩具等として、タッチスイッチが押圧操作されると、合成音を出力するロボット（本明細書においては、ぬいぐるみ状のものを含む）が数多く製品化されている。

【0003】

【発明が解決しようとする課題】

しかしながら、従来のロボットにおいては、タッチスイッチの押圧操作と、合成音との関係が固定であり、ユーザが飽きてしまう問題があった。

【0004】

本発明は、このような状況に鑑みてなされたものであり、エンタテインメント性の高いロボットを提供すること等ができるようにするものである。

【0005】

【課題を解決するための手段】

本発明の音声処理装置は、音声进行处理する音声処理手段と、ロボットの状態に基づいて、音声処理手段による音声処理を制御する制御手段とを備えることを特徴とする。

【0006】

制御手段には、ロボットの行動、感情、または本能の状態に基づいて、音声処理を制御させることができる。

【0007】

音声処理手段は、音声合成処理を行い、合成音を出力する音声合成手段で構成することができ、制御手段には、ロボットの状態に基づいて、音声合成手段による音声合成処理を制御させることができる。

【0008】

制御手段には、音声合成手段が出力する合成音の音韻情報または韻律情報を制御させることができる。

【0009】

また、制御手段には、音声合成手段が出力する合成音の発話速度または音量を制御させることができる。

【0010】

音声処理手段には、入力された音声の韻律情報若しくは音韻情報を抽出させることができ、この場合、ロボットの感情の状態を、韻律情報若しくは音韻情報に基づいて変更し、またはロボットには、韻律情報若しくは音韻情報に対応した行動をとらせることができる。

【0011】

音声処理手段は、入力された音声を認識する音声認識手段で構成することができ、ロボットには、音声認識手段が出力する音声認識結果の信頼性に対応する行動をとらせ、またはロボットの感情の状態を、信頼性に基づいて変更することができる。

【0012】

制御手段には、ロボットが行っている行動を認識させ、その行動に対する負荷に基づいて、音声処理手段による音声処理を制御させることができる。

【0013】

ロボットには、音声処理手段による音声処理に割り当て可能なリソースに対応する行動をとらせることができる。

【0014】

本発明の音声処理方法は、音声进行处理する音声処理ステップと、ロボットの状態に基づいて、音声処理ステップにおける音声処理を制御する制御ステップとを備えることを特徴とする。

【0015】

本発明の記録媒体は、音声进行处理する音声処理ステップと、ロボットの状態に基づいて、音声処理ステップにおける音声処理を制御する制御ステップとを備えるプログラムが記録されていることを特徴とする。

【0016】

本発明の音声処理装置および音声処理方法、並びに記録媒体においては、ロボットの状態に基づいて、音声処理が制御される。

【 0 0 1 7 】

## 【発明の実施の形態】

図 1 は、本発明を適用したロボットの一実施の形態の外観構成例を示しており、図 2 は、その電氣的構成例を示している。

【 0 0 1 8 】

本実施の形態では、ロボットは、犬形状のものとされており、胴体部ユニット 2 の前後左右に、それぞれ脚部ユニット 3 A、3 B、3 C、3 D が連結されるとともに、胴体部ユニット 2 の前端部と後端部に、それぞれ頭部ユニット 4 と尻尾部ユニット 5 が連結されることにより構成されている。

【 0 0 1 9 】

尻尾部ユニット 5 は、胴体部ユニット 2 の上面に設けられたベース部 5 B から、2 自由度をもって湾曲または揺動自在に引き出されている。

【 0 0 2 0 】

胴体部ユニット 2 には、ロボット全体の制御を行うコントローラ 1 0、ロボットの動力源となるバッテリー 1 1、並びにバッテリーセンサ 1 2 および熱センサ 1 3 からなる内部センサ部 1 4 などが収納されている。

【 0 0 2 1 】

頭部ユニット 4 には、「耳」に相当するマイク（マイクロフォン）1 5、「目」に相当する CCD (Charge Coupled Device) カメラ 1 6、触覚に相当するタッチセンサ 1 7、「口」に相当するスピーカ 1 8 などが、それぞれ所定位置に配設されている。

【 0 0 2 2 】

脚部ユニット 3 A 乃至 3 D それぞれの関節部分や、脚部ユニット 3 A 乃至 3 D それぞれと胴体部ユニット 2 の連結部分、頭部ユニット 4 と胴体部ユニット 2 の連結部分、並びに尻尾部ユニット 5 と胴体部ユニット 2 の連結部分などには、図 2 に示すように、それぞれアクチュエータ  $3 A A_1$  乃至  $3 A A_K$ 、 $3 B A_1$  乃至  $3 B A_K$ 、 $3 C A_1$  乃至  $3 C A_K$ 、 $3 D A_1$  乃至  $3 D A_K$ 、 $4 A_1$  乃至  $4 A_L$ 、 $5 A_1$  および  $5 A_2$  が配設されている。

【 0 0 2 3 】



頭部ユニット4におけるマイク15は、ユーザからの発話を含む周囲の音声（音）を集音し、得られた音声信号を、コントローラ10に送出する。CCDカメラ16は、周囲の状況を撮像し、得られた画像信号を、コントローラ10に送出する。

【0024】

タッチセンサ17は、例えば、頭部ユニット4の上部に設けられており、ユーザからの「なでる」や「たたく」といった物理的な働きかけにより受けた圧力を検出し、その検出結果を圧力検出信号としてコントローラ10に送出する。

【0025】

胴体部ユニット2におけるバッテリーセンサ12は、バッテリー11の残量を検出し、その検出結果を、バッテリー残量検出信号としてコントローラ10に送出する。熱センサ13は、ロボット内部の熱を検出し、その検出結果を、熱検出信号としてコントローラ10に送出する。

【0026】

コントローラ10は、CPU(Central Processing Unit)10Aやメモリ10B等を内蔵しており、CPU10Aにおいて、メモリ10Bに記憶された制御プログラムが実行されることにより、各種の処理を行う。

【0027】

即ち、コントローラ10は、マイク15や、CCDカメラ16、タッチセンサ17、バッテリーセンサ12、熱センサ13から与えられる音声信号、画像信号、圧力検出信号、バッテリー残量検出信号、熱検出信号に基づいて、周囲の状況や、ユーザからの指令、ユーザからの働きかけなどの有無を判断する。

【0028】

さらに、コントローラ10は、この判断結果等に基づいて、続く行動を決定し、その決定結果に基づいて、アクチュエータ $3A_1$ 乃至 $3A_K$ 、 $3B_1$ 乃至 $3B_K$ 、 $3C_1$ 乃至 $3C_K$ 、 $3D_1$ 乃至 $3D_K$ 、 $4A_1$ 乃至 $4A_L$ 、 $5A_1$ 、 $5A_2$ のうちの必要なものを駆動させ、これにより、頭部ユニット4を上下左右に振らせたり、尻尾部ユニット5を動かしたり、各脚部ユニット3A乃至3Dを駆動して、ロボットを歩行させるなどの行動を行わせる。

## 【 0 0 2 9 】

また、コントローラ 1 0 は、必要に応じて、合成音を生成し、スピーカ 1 8 に供給して出力させたり、ロボットの「目」の位置に設けられた図示しない L E D (Light Emitting Diode) を点灯、消灯または点滅させる。

## 【 0 0 3 0 】

以上のようにして、ロボットは、周囲の状況等に基づいて自律的に行動をとるようになっている。

## 【 0 0 3 1 】

次に、図 3 は、図 2 のコントローラ 1 0 の機能的構成例を示している。なお、図 3 に示す機能的構成は、C P U 1 0 A が、メモリ 1 0 B に記憶された制御プログラムを実行することで実現されるようになっている。

## 【 0 0 3 2 】

コントローラ 1 0 は、特定の外部状態を認識するセンサ入力処理部 5 0、センサ入力処理部 5 0 の認識結果を累積して、感情および本能の状態を表現する感情／本能モデル部 5 1、センサ入力処理部 5 0 の認識結果等に基づいて、続く行動を決定する行動決定機構部 5 2、行動決定機構部 5 2 の決定結果に基づいて、実際にロボットに行動を起こさせる姿勢遷移機構部 5 3、各アクチュエータ 3 A A<sub>1</sub>乃至 5 A<sub>1</sub>および 5 A<sub>2</sub>を駆動制御する制御機構部 5 4、並びに合成音を生成する音声合成部 5 5 から構成されている。

## 【 0 0 3 3 】

センサ入力処理部 5 0 は、マイク 1 5 や、C C D カメラ 1 6、タッチセンサ 1 7 等から与えられる音声信号、画像信号、圧力検出信号等に基づいて、特定の外部状態や、ユーザからの特定の働きかけ、ユーザからの指示等を認識し、その認識結果を表す状態認識情報を、感情／本能モデル部 5 1 および行動決定機構部 5 2 に通知する。

## 【 0 0 3 4 】

即ち、センサ入力処理部 5 0 は、音声認識部 5 0 A を有しており、音声認識部 5 0 A は、行動決定機構部 5 2 からの制御にしたがい、マイク 1 5 から与えられる音声信号を用いて、感情／本能モデル部 5 1 や行動決定機構部 5 2 から得られ

る情報を、必要に応じて考慮しながら、音声認識を行う。そして、音声認識部 5 0 A は、その音声認識結果としての、例えば、「歩け」、「伏せ」、「ボールを追いかける」等の指令その他を、状態認識情報として、感情／本能モデル部 5 1 および行動決定機構部 5 2 に通知する。

## 【 0 0 3 5 】

また、センサ入力処理部 5 0 は、画像認識部 5 0 B を有しており、画像認識部 5 0 B は、CCD カメラ 1 6 から与えられる画像信号を用いて、画像認識処理を行う。そして、画像認識部 5 0 B は、その処理の結果、例えば、「赤い丸いもの」や、「地面に対して垂直なかつ所定高さ以上の平面」等を検出したときには、「ボールがある」や、「壁がある」等の画像認識結果を、状態認識情報として、感情／本能モデル部 5 1 および行動決定機構部 5 2 に通知する。

## 【 0 0 3 6 】

さらに、センサ入力処理部 5 0 は、圧力処理部 5 0 C を有しており、圧力処理部 5 0 C は、タッチセンサ 1 7 から与えられる圧力検出信号を処理する。そして、圧力処理部 5 0 C は、その処理の結果、所定の閾値以上で、かつ短時間の圧力を検出したときには、「たたかれた（しかられた）」と認識し、所定の閾値未満で、かつ長時間の圧力を検出したときには、「なでられた（ほめられた）」と認識して、その認識結果を、状態認識情報として、感情／本能モデル部 5 1 および行動決定機構部 5 2 に通知する。

## 【 0 0 3 7 】

感情／本能モデル部 5 1 は、図 4 に示すような、ロボットの感情と本能の状態を表現する感情モデルと本能モデルをそれぞれ管理している。ここで、感情モデルおよび本能モデルは、図 3 のメモリ 1 0 B に記憶されている。

## 【 0 0 3 8 】

感情モデルは、例えば、3 つの感情ユニット 6 0 A, 6 0 B, 6 0 C で構成され、これらの感情ユニット 6 0 A 乃至 6 0 D は、「うれしさ」、「悲しさ」、「怒り」の感情の状態（度合い）を、例えば、0 乃至 1 0 0 の範囲の値によってそれぞれ表し、センサ入力処理部 5 0 からの状態認識情報や時間経過等に基づいて、その値を変化させる。

## 【0039】

なお、感情モデルには、「うれしさ」、「悲しさ」、「怒り」の他、「楽しさ」に対応する感情ユニットを設けることも可能である。

## 【0040】

本能モデルは、例えば、3つの本能ユニット61A、61B、61Cで構成され、これらの本能ユニット61A乃至61Cは、「食欲」、「睡眠欲」、「運動欲」という本能による欲求の状態（度合い）を、例えば、0乃至100の範囲の値によってそれぞれ表し、センサ入力処理部50からの状態認識情報や時間経過等に基づいて、その値を変化させる。

## 【0041】

感情／本能モデル部51は、上述のようにして変化する感情ユニット60A乃至60Cの値で表される感情の状態、および本能ユニット61A乃至61Cの値で表される本能の状態を、感情／本能状態情報として、センサ入力処理部50、行動決定機構部52、および音声合成部55に送出する。

## 【0042】

ここで、感情／本能モデル部51では、感情モデルを構成する感情ユニット60A乃至60Cどうしが、相互抑制的または相互刺激的に結合されており、これにより、結合している感情ユニットのうちの、ある1つの感情ユニットの値が変化すると、これに応じて、他の感情ユニットの値が変化し、これにより、自然な感情の変化が実現されるようになっている。

## 【0043】

即ち、例えば、図5（A）に示すように、感情モデルについては、「うれしさ」を表す感情ユニット60Aと、「悲しさ」を表す感情ユニット60Bとが相互抑制的に結合されており、ユーザに誉められたときには、まず、「うれしさ」の感情ユニット60Aの値が大きくなる。さらに、この場合、感情／本能モデル部51に対しては、「悲しさ」の感情ユニット60Bの値を変化させる状態認識情報が供給されていなくても、「うれしさ」の感情ユニット60Aの値が大きくなることに応じて、「悲しさ」の感情ユニット60Bの値が低下する。逆に、「悲しさ」の感情ユニット60Bの値が大きくなると、そのことに応じて、「うれし

さ」の感情ユニット60Aの値が低下する。

【0044】

また、「悲しさ」の感情ユニット60Bと、「怒り」の感情ユニット60Cとは、相互刺激的に結合されており、ユーザに叩かれたときには、まず、「怒り」の感情ユニット60Cの値が大きくなる。さらに、この場合、感情／本能モデル部51に対しては、「悲しさ」の感情ユニット60Bの値を変化させるような状態認識情報が供給されていなくても、「怒り」の感情ユニット60Cの値が大きくなることに応じて、「悲しさ」感情ユニット60Bの値が増大する。逆に、「悲しさ」の感情ユニット60Bの値が大きくなると、そのことに応じて、「怒り」の感情ユニット60Cの値が増大する。

【0045】

さらに、感情／本能モデル部51では、本能モデルを構成する感情ユニット61A乃至61Cどうしも、上述の感情モデルにおける場合と同様に、相互抑制的または相互刺激的に結合されており、結合している本能ユニットのうちの、ある1つの本能ユニットの値が変化すると、これに応じて、他の本能ユニットの値が変化し、これにより、自然な本能の変化が実現されるようになっている。

【0046】

また、感情／本能モデル部51には、センサ入力処理部50から状態認識情報が供給される他、行動決定機構部52から、ロボットの現在または過去の行動、具体的には、例えば、「長時間歩いた」などの行動の内容を示す行動情報が供給されるようになっており、同一の状態認識情報が与えられても、行動情報が示すロボットの行動に応じて、異なる感情／本能状態情報を生成するようになっている。

【0047】

即ち、例えば、図5(B)に示すように、感情モデルについては、各感情ユニット60A乃至60Cの前段に、行動情報および状態認識情報に基づいて、感情ユニット60A乃至60Cの値を増減させるための値情報を生成する強度増減関数65A乃至65Cがそれぞれ設けられており、強度増減関数65A乃至65Cから出力される値情報に応じて、感情ユニット60A乃至60Cの値が、それぞ

れ増減される。

【0048】

その結果、例えば、ロボットが、ユーザに挨拶をし、ユーザに頭を撫でられた場合には、ユーザに挨拶をしたという行動情報と、頭を撫でられたという状態認識情報とが、強度増減関数 6 5 A に与えられるが、この場合には、感情／本能モデル部 5 1 では、「うれしさ」の感情ユニット 6 0 A の値が増加される。

【0049】

一方、ロボットが、何らかの仕事を実行中に頭を撫でられた場合には、仕事を実行中であるという行動情報と、頭を撫でられたという状態認識情報とが、強度増減関数 6 5 A に与えられるが、この場合には、感情／本能モデル部 5 1 では、「うれしさ」の感情ユニット 6 0 A の値は変化されない。

【0050】

このように、感情／本能モデル部 5 1 は、状態認識情報だけでなく、現在または過去のロボットの行動を示す行動情報も参照しながら、感情ユニット 6 0 A 乃至 6 0 C の値を設定する。これにより、例えば、何らかのタスクを実行中に、ユーザが、いたずらするつもりで頭を撫でたときに、「うれしさ」の感情ユニット 6 0 A の値を増加させるような、不自然な感情の変化が生じることを回避することができる。

【0051】

なお、感情／本能モデル部 5 1 は、本能モデルを構成する本能ユニット 6 1 A 乃至 6 1 C についても、感情モデルにおける場合と同様に、状態認識情報および行動情報の両方に基づいて、それぞれの値を増減させるようになっている。

【0052】

ここで、強度増減関数 6 5 A 乃至 6 5 C は、状態認識情報および行動情報を入力として、あらかじめ設定されているパラメータに応じて、感情ユニット 6 0 A 乃至 6 0 C の値を変更するための値情報を生成して出力する関数であり、そのパラメータを、ロボットごとに異なる値に設定することにより、例えば、怒りっぽい性格のロボットや、明るい性格のロボットのように、ロボットに個性を持たせることができる。

## 【0 0 5 3】

図 3 に戻り、行動決定機構部 5 2 は、センサ入力処理部 5 0 からの状態認識情報や、感情／本能モデル部 5 1 からの感情／本能状態情報、時間経過等に基づいて、次の行動を決定し、決定された行動の内容を、行動指令情報として、姿勢遷移機構部 5 3 に送出する。

## 【0 0 5 4】

即ち、行動決定機構部 5 2 は、図 6 に示すように、ロボットがとり得る行動をステート（状態）(state)に対応させた有限オートマトンを、ロボットの行動を規定する行動モデルとして管理しており、この行動モデルとしての有限オートマトンにおけるステートを、センサ入力処理部 5 0 からの状態認識情報や、感情／本能モデル部 5 1 における感情モデルおよび本能モデルの値、時間経過等に基づいて遷移させ、遷移後のステートに対応する行動を、次にとるべき行動として決定する。

## 【0 0 5 5】

具体的には、例えば、図 6 において、ステート S T 3 が「立っている」という行動を、ステート S T 4 が「寝ている」という行動を、ステート S T 5 が「ボールを追いかけている」という行動を、それぞれ表しているとする。いま、例えば、「ボールを追いかけている」というステート S T 5 において、「ボールが見えなくなった」という状態認識情報が供給されると、ステート S T 5 から S T 3 に遷移し、その結果、ステート S T 3 に対応する「立っている」という行動を、次にとることが決定される。また、例えば、「寝ている」というステート S T 4 において、「起きろ」という状態認識情報が供給されると、ステート S T 4 から S T 3 に遷移し、その結果、やはり、ステート S T 3 に対応する「立っている」という行動を、次にとることが決定される。

## 【0 0 5 6】

ここで、行動決定機構部 5 2 は、所定のトリガ(trigger)があったことを検出すると、ステートを遷移させる。即ち、行動決定機構部 5 2 は、例えば、現在のステートに対応する行動を実行している時間が所定時間に達したときや、特定の状態認識情報を受信したとき、感情／本能モデル部 5 1 から供給される感情／本

能状態情報が示す感情の状態の値（感情ユニット60A乃至60Cの値）、あるいは本能の状態の値（本能ユニット61A乃至61Cの値）が所定の閾値以下または以上であるとき等に、ステートを遷移させる。

## 【0057】

なお、行動決定機構部52は、上述したように、センサ入力処理部50からの状態認識情報だけでなく、感情／本能モデル部51における感情モデルおよび本能モデルの値等にも基づいて、図6の有限オートマトンにおけるステートを遷移させることから、同一の状態認識情報が入力されても、感情モデルや本能モデルの値（感情／本能状態情報）によっては、ステートの遷移先は異なるものとなる。

## 【0058】

その結果、行動決定機構部52は、例えば、感情／本能状態情報が、「怒っていない」こと、および「お腹がすいていない」ことを表している場合において、状態認識情報が、「目の前に手のひらが差し出された」ことを表しているときには、目の前に手のひらが差し出されたことに応じて、「お手」という行動をとらせる行動指令情報を生成し、これを、姿勢遷移機構部53に送出する。

## 【0059】

また、行動決定機構部52は、例えば、感情／本能状態情報が、「怒っていない」こと、および「お腹がすいている」ことを表している場合において、状態認識情報が、「目の前に手のひらが差し出された」ことを表しているときには、目の前に手のひらが差し出されたことに応じて、「手のひらをぺろぺろなめる」ような行動を行わせるための行動指令情報を生成し、これを、姿勢遷移機構部53に送出する。

## 【0060】

また、行動決定機構部52は、例えば、感情／本能状態情報が、「怒っている」ことを表している場合において、状態認識情報が、「目の前に手のひらが差し出された」ことを表しているときには、感情／本能状態情報が、「お腹がすいている」ことを表していても、また、「お腹がすいていない」ことを表していても、「ぷいと横を向く」ような行動を行わせるための行動指令情報を生成し、これ



を、姿勢遷移機構部 5 3 に送出する。

【0 0 6 1】

なお、行動決定機構部 5 2 には、感情／本能モデル部 5 1 から供給される感情／本能状態情報が示す感情や本能の状態に基づいて、遷移先のステートに対応する行動のパラメータとしての、例えば、歩行の速度や、手足を動かす際の動きの大きさおよび速度などを決定させることができ、この場合、それらのパラメータを含む行動指令情報が、姿勢遷移機構部 5 3 に送出される。

【0 0 6 2】

また、行動決定機構部 5 2 では、上述したように、ロボットの頭部や手足等を動作させる行動指令情報の他、ロボットに発話を行わせる行動指令情報や、ロボットに音声認識を行わせる行動指令情報も生成される。ロボットに発話を行わせる行動指令情報は、音声合成部 5 5 に供給されるようになっており、音声合成部 5 5 に供給される行動指令情報には、音声合成部 5 5 に生成させる合成音に対応するテキスト等が含まれる。そして、音声合成部 5 5 は、行動決定部 5 2 から行動指令情報を受信すると、その行動指令情報に含まれるテキストに基づき、感情／本能モデル部 5 1 で管理されている感情の状態や本能の状態を加味しながら、合成音を生成し、スピーカ 1 8 に供給して出力させる。また、ロボットに音声認識を行わせる行動指令情報は、センサ入力処理部 5 0 の音声認識部 5 0 A に供給されるようになっており、音声認識部 5 0 A は、そのような行動指令情報を受信すると、音声認識処理を行う。

【0 0 6 3】

さらに、行動決定機構部 5 2 は、感情／本能モデル部 5 1 に供給するのと同じの行動情報を、センサ入力処理部 5 0 および音声合成部 5 5 に供給するようになっている。そして、センサ入力処理部 5 0 の音声認識部 5 0 A と、音声合成部 5 5 では、行動決定部 5 2 からの行動情報を加味して、音声認識と音声合成がそれぞれ行われる。この点については、後述する。

【0 0 6 4】

姿勢遷移機構部 5 3 は、行動決定機構部 5 2 から供給される行動指令情報に基づいて、ロボットの姿勢を、現在の姿勢から次の姿勢に遷移させるための姿勢遷

移情報を生成し、これを制御機構部 5 4 に送出する。

【0 0 6 5】

ここで、現在の姿勢から次に遷移可能な姿勢は、例えば、胴体や手や足の形状、重さ、各部の結合状態のようなロボットの物理的形状と、関節が曲がる方向や角度のようなアクチュエータ  $3 A A_1$  乃至  $5 A_1$  および  $5 A_2$  の機構とによって決定される。

【0 0 6 6】

また、次の姿勢としては、現在の姿勢から直接遷移可能な姿勢と、直接には遷移できない姿勢とがある。例えば、4 本足のロボットは、手足を大きく投げ出して寝転んでいる状態から、伏せた状態へ直接遷移することはできるが、立った状態へ直接遷移することはできず、一旦、手足を胴体近くに引き寄せて伏せた姿勢になり、それから立ち上がるという 2 段階の動作が必要である。また、安全に実行できない姿勢も存在する。例えば、4 本足のロボットは、その 4 本足で立っている姿勢から、両前足を挙げてバンザイをしようとする、簡単に転倒してしまう。

【0 0 6 7】

このため、姿勢遷移機構部 5 3 は、直接遷移可能な姿勢をあらかじめ登録しておき、行動決定機構部 5 2 から供給される行動指令情報が、直接遷移可能な姿勢を示す場合には、その行動指令情報を、そのまま姿勢遷移情報として、制御機構部 5 4 に送出する。一方、行動指令情報が、直接遷移不可能な姿勢を示す場合には、姿勢遷移機構部 5 3 は、遷移可能な他の姿勢に一旦遷移した後に、目的の姿勢まで遷移させるような姿勢遷移情報を生成し、制御機構部 5 4 に送出する。これによりロボットが、遷移不可能な姿勢を無理に実行しようとする事態や、転倒するような事態を回避することができるようになっている。

【0 0 6 8】

即ち、姿勢遷移機構部 5 3 は、例えば、図 7 に示すように、ロボットがとり得る姿勢をノード NODE 1 乃至 NODE 5 として表現するとともに、遷移可能な 2 つの姿勢に対応するノードどうしの間を、有向アーク ARC 1 乃至 ARC 1 0 で結合した有向グラフを記憶しており、この有向グラフに基づいて、上述したよ

うな姿勢遷移情報を生成する。

【0069】

具体的には、姿勢遷移機構部53は、行動決定機構部52から行動指令情報が供給されると、現在の姿勢に対応したノードNODEと、行動指令情報が示す次を取るべき姿勢に対応するノードNODEとを結ぶように、有向アークARCの向きに従いながら、現在のノードNODEから次のノードNODEに至る経路を探索し、探索した経路上にあるノードNODEに対応する姿勢を順番にとっていくように指示する姿勢遷移情報を生成する。

【0070】

その結果、姿勢遷移機構部53は、例えば、現在の姿勢が「ふせる」という姿勢を示すノードNODE2にある場合において、「すわれ」という行動指令情報が供給されると、有向グラフにおいて、「ふせる」という姿勢を示すノードNODE2から、「すわる」という姿勢を示すノードNODE5へは、直接遷移可能であることから、「すわる」に対応する姿勢遷移情報を生成して、制御機構部54に与える。

【0071】

また、姿勢遷移機構部53は、現在の姿勢が「ふせる」という姿勢を示すノードNODE2にある場合において、「歩け」という行動指令情報が供給されると、有向グラフにおいて、「ふせる」というノードNODE2から、「あるく」というノードNODE4に至る経路を探索する。この場合、「ふせる」に対応するノードNODE2、「たつ」に対応するNODE3、「あるく」に対応するNODE4の経路が得られるから、姿勢遷移機構部53は、「たつ」、「あるく」という順番の姿勢遷移情報を生成し、制御機構部54に送出する。

【0072】

制御機構部54は、姿勢遷移機構部53からの姿勢遷移情報にしたがって、アクチュエータ3AA<sub>1</sub>乃至5A<sub>1</sub>および5A<sub>2</sub>を駆動するための制御信号を生成し、これを、アクチュエータ3AA<sub>1</sub>乃至5A<sub>1</sub>および5A<sub>2</sub>に送出する。これにより、アクチュエータ3AA<sub>1</sub>乃至5A<sub>1</sub>および5A<sub>2</sub>は、制御信号にしたがって駆動し、ロボットは、自律的に行動を起こす。

## 【0073】

次に、図8は、図3の音声認識部50Aの構成例を示している。

## 【0074】

マイク15からの音声信号は、AD(Analog Digital)変換部21に供給される。AD変換部21では、マイク15からのアナログ信号である音声信号がサンプリング、量子化され、ディジタル信号である音声データにA/D変換される。この音声データは、特徴抽出部22に供給される。

## 【0075】

特徴抽出部22は、そこに入力される音声データについて、適当なフレームごとに、例えば、MFCC(Mel Frequency Cepstrum Coefficient)分析を行い、その分析結果を、特徴パラメータ(特徴ベクトル)として、マッチング部23に出力する。なお、特徴抽出部22では、その他、例えば、線形予測係数、ケプストラム係数、線スペクトル対、所定の周波数帯域ごとのパワー(フィルタバンクの出力)等を、特徴パラメータとして抽出することが可能である。

## 【0076】

また、特徴抽出部22は、そこに入力される音声データから韻律情報を抽出する。即ち、特徴抽出部22は、音声データを対象に、例えば、自己相関分析を行うことで、マイク15に入力された音声のピッチ周波数や、パワー(大きさ)、イントネーションに関する情報等の韻律情報を抽出する。

## 【0077】

マッチング部23は、特徴抽出部22からの特徴パラメータを用いて、音響モデル記憶部24、辞書記憶部25、および文法記憶部26を必要に応じて参照しながら、マイク15に入力された音声(入力音声)を、例えば、連続分布HMM(Hidden Markov Model)法に基づいて音声認識する。

## 【0078】

即ち、音響モデル記憶部24は、音声認識する音声の言語における個々の音素や音節などの音響的な特徴を表す音響モデルを記憶している。ここでは、連続分布HMM法に基づいて音声認識を行うので、音響モデルとしては、HMM(Hidden Markov Model)が用いられる。辞書記憶部25は、認識対象の各単語について

、その発音に関する情報（音韻情報）が記述された単語辞書を記憶している。文法記憶部 26 は、辞書記憶部 35 の単語辞書に登録されている各単語が、どのように連鎖する（つながる）かを記述した文法規則を記憶している。ここで、文法規則としては、例えば、文脈自由文法（CFG）や、統計的な単語連鎖確率（N-gram）などに基づく規則を用いることができる。

## 【0079】

マッチング部 23 は、辞書記憶部 25 の単語辞書を参照することにより、音響モデル記憶部 24 に記憶されている音響モデルを接続することで、単語の音響モデル（単語モデル）を構成する。さらに、マッチング部 23 は、幾つかの単語モデルを、文法記憶部 26 に記憶された文法規則を参照することにより接続し、そのようにして接続された単語モデルを用いて、特徴パラメータに基づき、連続分布 HMM 法によって、マイク 15 に入力された音声を認識する。即ち、マッチング部 23 は、特徴抽出部 22 が出力する時系列の特徴パラメータが観測されるスコア（尤度）が最も高い単語モデルの系列を検出し、その単語モデルの系列に対応する単語列の音韻情報（読み）を、音声の認識結果として出力する。

## 【0080】

即ち、マッチング部 23 は、接続された単語モデルに対応する単語列について、各特徴パラメータの出現確率を累積し、その累積値をスコアとして、そのスコアを最も高くする単語列の音韻情報を、音声認識結果として出力する。

## 【0081】

さらに、マッチング部 23 は、音声認識結果のスコアを、その音声認識結果の信頼性を表す信頼度情報として出力する。

## 【0082】

また、マッチング部 23 は、上述のようなスコア計算に伴って得られる、音声認識結果を構成する各音素や単語の継続時間長を検出し、マイク 15 に入力された音声の韻律情報として出力する。

## 【0083】

以上のようにして出力される、マイク 15 に入力された音声の認識結果、韻律情報、信頼度情報は、状態認識情報として、感情／本能モデル部 51 および行動

決定機構部 52 に出力される。

【0084】

以上のように構成される音声認識部 50A では、感情／本能モデル部 51 で管理されているロボットの感情や本能の状態に基づいて、音声認識処理が制御される。即ち、感情／本能モデル部 51 で管理されているロボットの感情や本能の状態は、特徴抽出部 22 およびマッチング部 23 に供給されるようになっており、特徴抽出部 22 およびマッチング部 23 は、そこに供給されるロボットの感情や本能の状態に基づいて、処理内容を変更するようになっている。

【0085】

具体的には、図 9 のフローチャートに示すように、行動決定機構部 52 から、音声認識処理を指示する行動指令情報が送信されてくると、ステップ S1 において、その行動指令情報が受信され、音声認識部 50A を構成する各ブロックがアクティブ状態にされる。これにより、音声認識部 50A は、マイク 15 に入力された音声を受け付けることが可能な状態とされる。

【0086】

なお、音声認識部 50A を構成する各ブロックは、常時、アクティブ状態しておくことが可能である。この場合、例えば、感情／本能モデル部 51 で管理されているロボットの感情や本能の状態が変化することにより、音声認識部 50A において、図 9 のステップ S2 以降の処理を開始するようになることが可能である。

【0087】

その後、特徴抽出部 22 およびマッチング部 23 は、ステップ S2 において、感情／本能モデル部 51 を参照することで、ロボットの感情や本能の状態を認識し、ステップ S3 に進む。ステップ S3 では、マッチング部 23 は、感情や本能の状態に基づいて、上述のスコア計算（マッチング）に用いる単語辞書を設定する。

【0088】

即ち、ここでは、辞書記憶部 25 は、音声認識の対象とする単語を、幾つかのカテゴリに分けて、そのカテゴリごとに単語が登録された複数の単語辞書を記憶しており、ステップ S3 では、ロボットの感情や本能の状態に基づいて、音声認

識に用いる単語辞書が設定される。

【0089】

具体的には、例えば、単語「お手」が登録されている単語辞書と、登録されていない単語辞書とが、辞書記憶部25に記憶されている場合において、ロボットの感情の状態が、「機嫌が良い」ことを表しているときには、単語「お手」が登録されている単語辞書が、音声認識に用いられるものとして設定される。また、ロボットの感情の状態が、「機嫌が悪い」ことを表しているときには、単語「お手」が登録されていない単語辞書が、音声認識に用いるものとして設定される。従って、ロボットの機嫌が良いときには、発話「お手」は音声認識され、その音声認識結果が、行動決定機構部52に供給されることにより、ロボットは、上述したようにして、発話「お手」に対応する行動をとる。一方、ロボットの機嫌が悪いときには、発話「お手」は音声認識されず（誤認識され）、その結果、ロボットは何の反応も起こさない（あるいは、発話「お手」に無関係な行動を起こす）。

【0090】

なお、ここでは、複数の単語辞書を用意しておき、ロボットの感情や本能の状態に基づいて、音声認識に用いる単語辞書を選択するようにしたが、その他、例えば、単語辞書は1つだけ用意しておき、ロボットの感情や本能の状態に基づいて、単語辞書の中から、音声認識の対象とする単語を選択するようにすることも可能である。

【0091】

ステップS3の処理後は、ステップS4に進み、特徴抽出部22およびマッチング部23は、ロボットの感情や本能の状態に基づいて、音声認識処理に用いるパラメータ（認識パラメータ）を設定する。

【0092】

即ち、特徴抽出部22およびマッチング部23は、例えば、ロボットの感情の状態が「怒っている」ことを表しているときや、ロボットの本能の状態が「眠い」ことを表しているときには、音声認識精度が劣化するように、認識パラメータを設定する。一方、例えば、ロボットの感情の状態が「機嫌が良い」ことを表し

ているときには、音声認識精度が向上するように、認識パラメータを設定する。

【0093】

ここで、音声認識精度に影響を与える認識パラメータとしては、例えば、音声区間の検出に用いる、マイク15に入力された音声と比較する閾値等がある。

【0094】

その後、ステップS5に進み、マイク15に入力された音声は、AD変換部21を介して、特徴抽出部22に取り込まれ、ステップS6に進む。ステップS6では、特徴抽出部22およびマッチング部23において、ステップS3およびS4で行われた設定の下、上述したような処理が行われることにより、マイク15に入力された音声は音声認識される。そして、ステップS7に進み、ステップS6の処理によって得られる音声認識結果としての音韻情報、韻律情報、信頼度情報が、状態認識情報として、感情／本能モデル部51および行動決定機構部52に出力され、処理を終了する。

【0095】

感情／本能モデル部51は、以上のような状態認識情報を、音声認識部50Aから受信すると、その状態認識情報に基づいて、図5で説明したようにして、感情モデルや本能モデルの値を変更し、これにより、ロボットの感情や本能の状態を変化させる。

【0096】

即ち、例えば、状態認識情報における音声認識結果としての音韻情報が「ばか」である場合には、感情／本能モデル部51は、「怒り」の感情ユニット60Cの値を大きくする。また、感情／本能モデル部51は、状態認識情報における韻律情報としてのピット周波数や、パワー、継続時間長に基づいて、強度増減関数65A乃至65Cが出力する値情報を変化させ、これにより、感情モデルや本能モデルの値を変更する。

【0097】

また、状態認識情報における信頼度情報が、音声認識結果の信頼性が低いことを表しているときには、感情／本能モデル部51は、例えば、「悲しさ」の感情ユニット60Bの値を大きくする。一方、状態認識情報における信頼度情報が、



音声認識結果の信頼性が高いことを表しているときには、感情／本能モデル部 51 は、例えば、「うれしさ」の感情ユニット 60A の値を大きくする。

## 【0098】

行動決定機構部 52 は、音声認識部 50A から状態認識情報を受信すると、その状態認識情報に基づいて、ロボットの次の行動を決定し、その行動を表す行動指令情報を生成する。

## 【0099】

即ち、行動決定機構部 52 は、例えば、上述したように、状態認識情報における音声認識結果の音韻情報に対応する行動をとることを決定する（例えば、音声認識結果が「お手」であれば、お手の行動をとることを決定する）。

## 【0100】

あるいは、また、行動決定機構部 52 は、状態認識情報における信頼度情報が、音声認識結果の信頼性が低いことを表しているときには、例えば、首をかしげるような、またはすまなさそうな行動をとることを決定する。また、行動決定機構部 52 は、状態認識情報における信頼度情報が、音声認識結果の信頼性が高いことを表しているとき、例えば、うなづくような行動をとることを決定する。この場合、ユーザに対して、ロボットにおける、ユーザの発話の理解の程度を示すことができる。

## 【0101】

次に、音声認識部 50A に対しては、上述したように、行動決定機構部 52 から、ロボットの現在または過去の行動の内容を示す行動情報が供給されるようになっており、音声認識部 50A では、この行動情報に基づいて、音声認識処理の制御を行うようにすることも可能である。即ち、行動決定機構部 52 が出力する行動情報を、特徴抽出部 22 やマッチング部 23 に供給し、特徴抽出部 22 やマッチング部 23 には、そこに供給される行動情報に基づいて、処理内容を変更させるようにすることが可能である。

## 【0102】

具体的には、図 10 のフローチャートに示すように、行動決定機構部 52 から、音声認識処理を指示する行動指令情報が送信されてくると、音声認識部 50A

では、ステップ S 1 1 において、図 9 のステップ S 1 における場合と同様に、その行動指令情報が受信され、音声認識部 5 0 A を構成する各ブロックがアクティブ状態にされる。

#### 【0103】

なお、上述したように、音声認識部 5 0 A を構成する各ブロックは、常時、アクティブ状態しておくことが可能であり、この場合、例えば、行動決定機構部 5 2 が出力する行動情報が変化すると、音声認識部 5 0 A において、図 1 0 のステップ S 1 2 以降の処理を開始するようにすることが可能である。

#### 【0104】

その後、特徴抽出部 2 2 およびマッチング部 2 3 は、ステップ S 1 2 において、行動決定機構部 5 2 が出力する行動情報を参照し、ステップ S 1 3 に進む。ステップ S 1 3 では、マッチング部 2 3 は、行動情報に基づいて、上述のスコア計算（マッチング）に用いる単語辞書を設定する。

#### 【0105】

即ち、例えば、行動情報が、現在の行動が「座っている」、あるいは「ねそべっている」ことを表している場合に、ユーザが、「お座り」といった発話を行うことは、ほとんどないと考えられる。そこで、行動情報が、現在の行動が「座っている」、あるいは「ねそべっている」ことを表している場合においては、マッチング部 2 5 は、単語「お座り」を、音声認識の対象から除外するように、辞書記憶部 2 5 における単語辞書を設定する。この場合、発話「お座り」は音声認識されないことになる。さらに、この場合、音声認識の対象とする単語が減少するので、処理の高速化、および認識精度の向上を図ることが可能となる。

#### 【0106】

ステップ S 1 3 の処理後は、ステップ S 1 4 に進み、特徴抽出部 2 2 およびマッチング部 2 3 は、行動情報に基づいて、音声認識処理に用いるパラメータ（認識パラメータ）を設定する。

#### 【0107】

即ち、特徴抽出部 2 2 およびマッチング部 2 3 は、例えば、行動情報が、「歩いている」ことを表している場合には、「座っている」ことや「伏せている」こ

と等を表している場合に比較して、認識パラメータを、処理速度よりも、精度を優先するように設定する。

## 【0108】

一方、例えば、行動情報が、「座っている」ことや「伏せている」こと等を表している場合には、「歩いている」ことを表している場合に比較して、認識パラメータを、精度よりも、処理速度を優先するように設定する。

## 【0109】

ロボットが歩いている場合には、座っている場合や、伏せている場合に比較して、アクチュエータ  $3A_{A_1}$  乃至  $5A_1$  および  $5A_2$  の駆動による雑音のレベルが高くなることから、その雑音の影響で、一般に、音声認識の精度が劣化する。そこで、ロボットが歩いている場合には、認識パラメータを、処理速度よりも、精度を優先するように設定することで、そのような雑音による音声認識精度の劣化を防止（低減）することが可能となる。

## 【0110】

一方、ロボットが、座っている場合や、伏せている場合には、上述のようなアクチュエータ  $3A_{A_1}$  乃至  $5A_1$  および  $5A_2$  の駆動による雑音は存在しないから、その雑音による音声認識精度の劣化もない。そこで、ロボットが、座っている場合や、伏せている場合には、認識パラメータを、精度よりも、処理速度を優先するように設定することで、ある程度の音声認識精度を維持しながら、音声認識の処理速度を向上させることが可能となる。

## 【0111】

ここで、音声認識の精度および処理速度に影響を与える認識パラメータとしては、例えば、マッチング部 23 において、スコア計算の対象とする範囲をビームサーチ法により制限する場合における仮説の範囲（ビームサーチする際のビーム幅）等がある。

## 【0112】

その後、ステップ S15 に進み、マイク 15 に入力された音声が入力部 21 を介して、特徴抽出部 22 に取り込まれ、ステップ S16 に進む。ステップ S16 では、特徴抽出部 22 およびマッチング部 23 において、ステップ S13

および S 14 で行われた設定の下、上述したような処理が行われることにより、マイク 15 に入力された音声は音声認識される。そして、ステップ S 17 に進み、ステップ S 16 の処理によって得られる音声認識結果としての音韻情報、韻律情報、信頼度情報が、状態認識情報として、感情／本能モデル部 51 および行動決定機構部 52 に出力され、処理を終了する。

#### 【0113】

感情／本能モデル部 51 および行動決定機構部 52 は、以上のような状態認識情報を、音声認識部 50A から受信すると、その状態認識情報に基づいて、上述したように、感情モデルや本能モデルの値を変更するとともに、ロボットの次の行動を決定する。

#### 【0114】

なお、上述の場合には、ロボットが歩いているときに、アクチュエータ 3A<sub>1</sub>乃至 5A<sub>1</sub>および 5A<sub>2</sub>の駆動による雑音の影響によって、音声認識の精度が劣化することから、認識パラメータを、処理速度よりも、精度を優先するように設定するようにすることで、雑音による音声認識精度の劣化を防止するようにしたが、その他、ロボットが歩いているときには、ロボットを、一旦停止させて、音声認識を行うようにすることが可能であり、このようにすることによっても、音声認識の精度が劣化することを防止することが可能である。

#### 【0115】

次に、図 11 は、図 3 の音声合成部 55 の構成例を示している。

#### 【0116】

テキスト生成部 31 には、行動決定機構部 52 が出力する、音声合成の対象とするテキストを含む行動指令情報が供給されるようになっており、テキスト生成部 31 は、辞書記憶部 34 や解析用文法記憶部 35 を参照しながら、その行動指令情報に含まれるテキストを解析する。

#### 【0117】

即ち、辞書記憶部 34 には、各単語の品詞情報や、読み、アクセント等の情報が記述された単語辞書が記憶されており、また、解析用文法記憶部 35 には、辞書記憶部 34 の単語辞書に記述された単語について、単語連鎖に関する制約等の

解析用文法規則が記憶されている。そして、テキスト生成部 31 は、この単語辞書および解析用文法規則に基づいて、そこに入力されるテキストの形態素解析や構文解析等の解析を行い、後段の規則合成部 32 で行われる規則音声合成に必要な情報を抽出する。ここで、規則音声合成に必要な情報としては、例えば、ポーズの位置や、アクセントおよびイントネーションを制御するための情報その他の韻律情報や、各単語の発音等の音韻情報などがある。

## 【0118】

テキスト生成部 31 で得られた情報は、規則合成部 32 に供給され、規則合成部 32 では、音素片記憶部 36 を用いて、テキスト生成部 31 に入力されたテキストに対応する合成音の音声データ（デジタルデータ）が生成される。

## 【0119】

即ち、音素片記憶部 36 には、例えば、CV (Consonant, Vowel) や、VCV、CVC 等の形で音素片データが記憶されており、規則合成部 32 は、テキスト生成部 31 からの情報に基づいて、必要な音素片データを接続し、さらに、ポーズ、アクセント、イントネーション等を適切に付加することで、テキスト生成部 31 に入力されたテキストに対応する合成音の音声データを生成する。

## 【0120】

この音声データは、DA (Digital Analogue) 変換部 33 に供給され、そこで、アナログ信号としての音声信号に D/A 変換される。この音声信号は、スピーカ 18 に供給され、これにより、テキスト生成部 31 に入力されたテキストに対応する合成音が出力される。

## 【0121】

以上のように構成される音声合成部 55 には、行動決定機構部 52 から、音声合成の対象とするテキストを含む行動指令情報の他、感情／本能モデル部 51 から感情および本能の状態が供給されるととともに、行動決定機構部 52 から行動情報が供給されるようになっており、テキスト生成部 31 および規則合成部 32 は、これらの感情や本能の状態、行動情報を考慮して音声合成処理を行うようになっている。

## 【0122】

そこで、まず、図12のフローチャートを参照して、感情や本能の状態を考慮して行われる音声合成処理について説明する。

#### 【0123】

行動決定機構部52が、音声合成の対象とするテキストを含む行動指令情報を、音声合成部55に出力すると、テキスト生成部31は、ステップS21において、その行動指令情報を受信し、ステップS22に進む。ステップS22では、テキスト生成部31および規則合成部32において、感情／本能モデル部51を参照することで、ロボットの感情や本能の状態が認識され、ステップS23に進む。

#### 【0124】

ステップS23では、テキスト生成部31において、行動決定機構部52からの行動指令情報に含まれるテキストから、実際に合成音として出力するテキスト（以下、適宜、発話テキストという）を生成する際に用いる語彙（発話語彙）が、ロボットの感情や本能の状態に基づいて設定され、ステップS24に進む。ステップS24では、テキスト生成部31において、ステップS23で設定された発話語彙を用いて、行動指令情報に含まれるテキストに対応する発話テキストが生成される。

#### 【0125】

即ち、行動決定機構部52からの行動指令情報に含まれるテキストは、例えば、標準的な感情および本能の状態における発話を前提としたものとなっており、ステップS24では、そのテキストが、ロボットの感情や本能の状態を考慮して修正され、これにより、発話テキストが生成される。

#### 【0126】

具体的には、例えば、行動指令情報に含まれるテキストが、「何ですか？」である場合において、ロボットの感情の状態が「怒っている」ことを表しているときには、その怒りを表現する「何だよ！」が、発話テキストとして生成される。あるいは、また、例えば、行動指令情報に含まれるテキストが、「やめて下さい」である場合において、ロボットの感情の状態が「怒っている」ことを表しているときには、その怒りを表現する「やめろ！」が、発話テキストとして生成され

る。

【0127】

そして、ステップS25に進み、テキスト生成部31は、発話テキストを対象に、形態素解析や構文解析等のテキスト解析を行い、その発話テキストについて規則音声合成を行うのに必要な情報としての、ピッチ周波数や、パワー、継続時間長等の韻律情報を生成する。さらに、テキスト生成部31は、発話テキストを構成する各単語の発音等の音韻情報も生成する。ここで、ステップS25では、発話テキストの韻律情報として、標準的な韻律情報が生成される。

【0128】

その後、テキスト生成部31は、ステップS26において、ステップS25で設定した発話テキストの韻律情報を、ロボットの感情や本能の状態に基づいて修正し、これにより、発話テキストが合成音で出力されときの感情表現が高められる。

【0129】

ここで、感情と音声との関係に関しては、例えば、前川、「音声によるパラ言語情報の伝達：言語学の立場から」、日本音響学会、平成9年度秋季研究発表会講演論文集1-3-10、pp. 381-384、平成9年9月等に、その詳細が記載されている。

【0130】

テキスト生成部31で得られた発話テキストの音韻情報および韻律情報は、規則合成部32に供給され、規則合成部32では、ステップS27において、その音韻情報および韻律情報にしたがい、規則音声合成が行われることにより、発話テキストの合成音のデジタルデータが生成される。ここで、規則合成部32でも、規則音声合成の際、ロボットの感情や本能の状態に基づいて、その感情や本能の状態を適切に表現するように、合成音のポーズの位置や、アクセントの位置、イントネーション等の韻律が変更される。

【0131】

規則合成部32で得られた合成音のデジタルデータは、DA変換部33に供給される。DA変換部33では、ステップS28において、規則合成部32から

のデジタルデータがD/A変換され、スピーカ18に供給されて、処理を終了する。これにより、スピーカ18からは、発話テキストの合成音であって、ロボットの感情や本能の状態を反映した韻律を有するものが出力される。

#### 【0132】

次に、図13のフローチャートを参照して、行動情報を考慮して行われる音声合成処理について説明する。

#### 【0133】

行動決定機構部52が、音声合成の対象とするテキストを含む行動指令情報を、音声合成部55に出力すると、テキスト生成部31は、ステップS31において、その行動指令情報を受信し、ステップS32に進む。ステップS32では、テキスト生成部31および規則合成部32において、行動決定機構部52が出力する行動情報が参照され、これにより、ロボットの現在の行動が認識されて、ステップS33に進む。

#### 【0134】

ステップS33では、テキスト生成部31において、行動決定機構部52からの行動指令情報に含まれるテキストから、発話テキストを生成する際に用いる語彙（発話語彙）が、行動情報に基づいて設定され、その発話語彙を用いて、行動指令情報に含まれるテキストに対応する発話テキストが生成される。

#### 【0135】

そして、ステップS34に進み、テキスト生成部31は、発話テキストを対象に、形態素解析や構文解析等のテキスト解析を行い、その発話テキストについて規則音声合成を行うのに必要な情報としての、ピッチ周波数や、パワー、継続時間長等の韻律情報を生成する。さらに、テキスト生成部31は、発話テキストを構成する各単語の発音等の音韻情報も生成する。ここで、ステップS34でも、図12のステップS25における場合と同様に、発話テキストの韻律情報としては、標準的なものが生成される。

#### 【0136】

その後、テキスト生成部31は、ステップS35において、ステップS25で生成した発話テキストの韻律情報を、行動情報に基づいて修正する。



## 【0 1 3 7】

即ち、例えば、ロボットが歩いている場合には、上述したように、アクチュエータ  $3 A A_1$  乃至  $5 A_1$  および  $5 A_2$  の駆動による雑音が存在する。一方、ロボットが、座っている場合や、伏せている場合には、そのような雑音は存在しない。従って、ロボットが歩いている場合には、座っている場合や、伏せている場合に比較して、合成音が聞き取りにくくなる。

## 【0 1 3 8】

そこで、テキスト生成部 3 1 は、行動情報が、ロボットが歩いていることを表している場合には、合成音の発話速度を遅くしたり、パワーを大きくするように、韻律情報を修正し、合成音を聞き取りやすくする。

## 【0 1 3 9】

その他、ステップ S 3 5 では、例えば、行動情報が、寝ていることを表している場合と、立っていることを表している場合とで、ピッチ周波数が異なる値となるように、修正を行うことも可能である。

## 【0 1 4 0】

テキスト生成部 3 1 で得られた発話テキストの音韻情報および韻律情報は、規則合成部 3 2 に供給され、規則合成部 3 2 では、ステップ S 3 6 において、その音韻情報および韻律情報にしたがい、規則音声合成が行われることにより、発話テキストの合成音のデジタルデータが生成される。ここで、規則合成部 3 2 でも、規則音声合成の際、行動情報に基づいて、合成音のポーズの位置や、アクセントの位置、イントネーション等が、必要に応じて変更される。

## 【0 1 4 1】

規則合成部 3 2 で得られた合成音のデジタルデータは、D/A変換部 3 3 に供給される。D/A変換部 3 3 では、ステップ S 3 7 において、規則合成部 3 2 からのデジタルデータがD/A変換され、スピーカ 1 8 に供給されて、処理を終了する。

## 【0 1 4 2】

なお、以上のように、音声合成部 5 5 において、感情や本能の状態、行動情報を考慮した合成音を生成する場合においては、そのような合成音の出力と、ロボ

ットの行動とを、いわば同期させることが可能である。

【0143】

即ち、例えば、感情の状態が「怒っていない」ことを表している場合において、その感情の状態を考慮して、合成音「何ですか？」を出力する場合には、その合成音の出力に同期して、ロボットを振り向かせるようにすることが可能である。一方、例えば、感情の状態が「怒っている」ことを表している場合において、その感情の状態を考慮して、合成音「何だよ！」を出力する場合には、その合成音の出力に同期して、ロボットにそっぽを向かせるようにすることが可能である。

【0144】

また、合成音「何ですか？」を出力する場合には、ロボットに、通常の数値で行動させ、合成音「何だよ！」を出力する場合には、ロボットに、通常の数値より遅い数値で、いわばのらりくらりと不満げに行動させるようにすることが可能である。

【0145】

この場合、ユーザに対して、動きと合成音の両方で、感情を表現することができる。

【0146】

さらに、行動決定機構部52では、図6に示したような有限オートマトンで表される行動モデルに基づいて、次の行動が決定されるが、合成音として出力するテキストの内容は、図6の行動モデルのステートの遷移に対応付けておくことが可能である。

【0147】

即ち、例えば、行動「座る」に対応するステートから、行動「立つ」に対応するステートへの遷移には、テキスト「よっこいしょ」などを対応付けておくことが可能である。この場合、ロボットが、座っている姿勢から、立つ姿勢に移行するときに、その姿勢の移行に同期して、合成音「よっこいしょ」を出力することが可能となる。

【0148】

以上のように、ロボットの状態に基づいて、音声合成処理や音声認識処理を制御することで、エンタテインメント性の高いロボットを提供すること等が可能となる。

#### 【0149】

次に、図14は、図3のセンサ入力処理部50を構成する画像認識部50Bの構成例を示している。

#### 【0150】

CCDカメラ16が出力する画像信号は、AD変換部41に供給され、そこでA/D変換されることにより、デジタルの画像データとされる。このデジタル画像データは、画像処理部42に供給される。画像処理部42では、AD変換部41からの画像データに対して、例えば、DCT(Discrete Cosine Transform)等の所定の画像処理が施され、認識照合部43に供給される。

#### 【0151】

認識照合部43は、画像パターン記憶部44に記憶された複数の画像パターンそれぞれと、画像処理部42の出力との間の距離を計算し、その距離を最も小さくする画像パターンを検出する。そして、認識照合部43は、その検出した画像パターンに基づいて、CCDカメラ16で撮影された画像を認識し、その認識結果を、状態認識情報として、感情/本能モデル部51および行動決定機構部52に出力する。

#### 【0152】

ところで、図3のブロック図に示した構成は、上述したように、CPU10Aが制御プログラムを実行することで実現される。いま、例えば、音声認識部50Aを実現するために必要なリソースとして、CPU10Aのパワー（以下、適宜、CPUパワーという）だけを考えて、CPUパワーは、CPUA10Aとして採用するハードウェアによって一意に決まり、そのCPUパワーによって行うことのできる処理量（ある単位時間あたりの処理量）も一意に決まる。

#### 【0153】

一方、CPU10Aが行うべき処理の中には、音声認識処理よりも優先して行わなければならない処理（以下、適宜、優先処理という）があり、従って、優先

処理に対するCPU10Aの負荷が増えれば、音声認識処理に割り当てることのできるCPUパワーは少なくなる。

## 【0154】

即ち、優先処理に対するCPU10Aの負荷を $x\%$ で表すとともに、音声認識処理に割り当てることのできるCPUパワーを $y\%$ で表すと、 $x$ と $y$ との関係は、式 $x + y = 100\%$ で表され、図15に示すようになる。

## 【0155】

従って、優先処理に対する負荷が $0\%$ である場合には、音声認識処理には、 $100\%$ のCPUパワーを割り当てることのできる。また、優先処理に対する負荷が $S$  ( $0 < S < 100$ )  $\%$ である場合には、音声認識処理には、 $100 - S\%$ のCPUパワーを割り当てることのできる。そして、優先処理に対する負荷が $100\%$ である場合には、音声認識処理には、CPUパワーを割り当てることのできない。

## 【0156】

ここで、例えば、ロボットが歩いているときに、その「歩く」という行動を行わせるための処理（以下、適宜、歩行処理という）に割り当てるCPUパワーが不足すると、歩く速度が遅くなり、最悪の場合は停止する。このように、歩く速度が遅くなったり、停止することは、ユーザに違和感を感じさせるから、そのようなことが生じることは極力防止する必要がある、従って、ロボットが歩いている場合における歩行処理は、音声認識処理よりも優先して行わなければならない優先処理ということができる。

## 【0157】

即ち、現在行われている処理が、音声認識処理が行われることにより妨げられ、ロボットの行動がスムーズに行われなくなると、ユーザに違和感を感じさせることになる。従って、現在行われている処理は、基本的には、音声認識処理よりも優先して行わなければならない優先処理ということができ、音声認識処理は、現在行われている処理を妨げない範囲で行うべきである。

## 【0158】

そこで、行動決定機構部52は、ロボットが行っている行動を認識し、その行

動に対する負荷に基づいて、音声認識部 50A による音声認識処理を制御するようになっている。

【0159】

即ち、図 16 のフローチャートに示すように、行動決定機構部 52 は、ステップ S41 において、自身が管理している行動モデルに基づいて、ロボットが現在行っている行動を認識し、ステップ S42 に進む。ステップ S42 では、行動決定機構部 52 は、ステップ S41 で認識した現在の行動をそのまま続行させる（維持する）ための処理に対する負荷を認識する。

【0160】

ここで、現在の行動をそのまま続行させるための処理に対する負荷は、所定の計算によって求めることが可能である。また、負荷は、行動と、その行動に対応する処理を行うために予想される CPU パワーとを対応付けたテーブルをあらかじめ用意しておき、そのテーブルを参照することで求めることも可能である。なお、計算による場合よりも、テーブルによる場合の方が、処理量が少なくて済む。

【0161】

現在の行動をそのまま続行させるための処理に対する負荷を求めた後は、ステップ S43 に進み、行動決定機構部 52 は、その負荷に基づき、図 15 に示した関係から、音声認識処理に割り当て可能な CPU パワーを求める。さらに、行動決定機構部 52 は、その音声認識処理に割り当て可能な CPU パワーに基づき、音声認識処理に関する各種の制御を行い、ステップ S41 に戻り、以下、同様の処理を繰り返す。

【0162】

即ち、行動決定機構部 52 は、例えば、音声認識処理に割り当て可能な CPU パワーに基づき、音声認識処理に用いる単語辞書を変更する。具体的には、音声認識処理に対して、十分な CPU パワーを割り当てることができる場合には、多くの単語が登録されている単語辞書を、音声認識処理に用いるように、設定を行う。また、音声認識処理に対して、十分な CPU パワーを割り当てることができない場合には、少ない単語が登録されている単語辞書を、音声認識に用いるよう

に、設定を行う。

【0 1 6 3】

さらに、行動決定機構部 5 2 は、音声認識処理に対して、CPU パワーを、ほとんど割り当てることができない場合には、音声認識部 5 0 A をスリープ状態にする（音声認識処理を行わないようにする）。

【0 1 6 4】

また、行動決定機構部 5 2 は、音声認識処理に割り当て可能な CPU パワーに対応する行動を、ロボットに起こさせる。

【0 1 6 5】

即ち、音声認識処理に対して、ほとんど CPU パワーを割り当てることができない場合や、十分な CPU パワーを割り当てることができない場合には、音声認識処理が行われず、あるいは、音声認識精度や処理速度が劣化するから、ユーザに違和感を感じさせることがある。

【0 1 6 6】

そこで、行動決定機構部 5 2 は、音声認識処理に対して、CPU パワーを、ほとんど割り当てることができない場合や、十分な CPU パワーを割り当てることができない場合には、例えば、ロボットに、元気のない行動や、首をかしげるような行動をとらせ、これにより、ユーザに対して、音声認識が困難である旨を報知する。

【0 1 6 7】

また、行動決定機構部 5 2 は、音声認識処理に対して、十分な CPU パワーを割り当てることができる場合には、例えば、ロボットに、元気な行動やうなずくような行動をとらせ、これにより、ユーザに対して、音声認識が十分に可能である旨を報知する。

【0 1 6 8】

ここで、音声認識処理が可能であるかどうかは、ロボットに、上述のような行動をとらせることで、ユーザに報知する他、例えば、「ピーピーピー」や「ピョロピョロピョロ」等の特殊な音や、所定のメッセージの合成音を、スピーカ 1 8 から出力することで、ユーザに報知することも可能である。

【0 1 6 9】

また、ロボットが、液晶パネルを有する場合には、その液晶パネルに、所定のメッセージを表示することで、ユーザに、音声認識処理が可能かどうかを報知することが可能である。さらに、ロボットが、例えば、瞬きをする等の顔の表情を表すことのできる機構を有する場合には、その機構によって、顔の表情を変更することで、ユーザに、音声認識処理が可能かどうかを報知することが可能である。

【0 1 7 0】

なお、上述の場合においては、CPUパワーだけを対象としたが、音声認識処理に必要なその他のリソース（例えば、メモリ 1 0 B の空き容量等）をも対象とすることが可能である。

【0 1 7 1】

さらに、上述の場合には、音声認識部 5 0 A における音声認識処理と、他の処理との関係に注目して説明したが、その他、画像認識部 5 0 B における画像認識処理と他の処理との関係や、音声合成部 5 5 における音声合成処理と他の処理との関係等についても、同様のことがいえる。

【0 1 7 2】

以上、本発明を、エンターテイメント用のロボット（疑似ペットとしてのロボット）に適用した場合について説明したが、本発明は、これに限らず、例えば、産業用のロボット等の各種のロボットに広く適用することが可能である。

【0 1 7 3】

さらに、本実施の形態においては、上述した一連の処理を、CPU 1 0 A にプログラムを実行させることにより行うようにしたが、一連の処理は、それ専用のハードウェアによって行うことも可能である。

【0 1 7 4】

なお、プログラムは、あらかじめメモリ 1 0 B（図 2）に記憶させておく他、フロッピーディスク、CD-ROM (Compact Disc Read Only Memory)、MO (Magneto optical) ディスク、DVD (Digital Versatile Disc)、磁気ディスク、半導体メモリなどのリムーバブル記録媒体に、一時的あるいは永続的に格納（記録）しておく

ことができる。そして、このようなリムーバブル記録媒体を、いわゆるパッケージソフトウェアとして提供し、ロボット（メモリ 1 0 B）にインストールするようにすることができる。

【 0 1 7 5 】

また、プログラムは、リムーバブル記録媒体からインストールする他、ダウンロードサイトから、ディジタル衛星放送用の人工衛星を介して、無線で転送したり、LAN(Local Area Network)、インターネットといったネットワークを介して、有線で転送し、メモリ 1 0 B にインストールすることができる。

【 0 1 7 6 】

この場合、プログラムがバージョンアップされたとき等に、そのバージョンアップされたプログラムを、メモリ 1 0 B に、容易にインストールすることができる。

【 0 1 7 7 】

ここで、本明細書において、CPU 1 0 A に各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含むものである。

【 0 1 7 8 】

また、プログラムは、1 の CPU により処理されるものであっても良いし、複数の CPU によって分散処理されるものであっても良い。

【 0 1 7 9 】

【発明の効果】

本発明の音声処理装置および音声処理方法、並びに記録媒体によれば、ロボットの状態に基づいて、音声処理が制御される。従って、エンタテインメント性の高いロボットを提供すること等が可能となる。

【図面の簡単な説明】

【図 1】

本発明を適用したロボットの一実施の形態の外観構成例を示す斜視図である。

【図 2】



図 1 のロボットの内部構成例を示すブロック図である。

【図 3】

図 2 のコントローラ 1 0 の機能的構成例を示すブロック図である。

【図 4】

感情／本能モデルを示す図である。

【図 5】

感情／本能モデル部 5 1 における処理を説明するための図である。

【図 6】

行動モデルを示す図である。

【図 7】

姿勢遷移機構部 5 4 の処理を説明するための図である。

【図 8】

音声認識部 5 0 A の構成例を示すブロック図である。

【図 9】

音声認識部 5 0 A の処理を説明するためのフローチャートである。

【図 1 0】

音声認識部 5 0 A の処理を説明するためのフローチャートである。

【図 1 1】

音声合成部 5 5 の構成例を示すブロック図である。

【図 1 2】

音声合成部 5 5 の処理を説明するためのフローチャートである。

【図 1 3】

音声合成部 5 5 の処理を説明するためのフローチャートである。

【図 1 4】

画像認識部 5 0 B の構成例を示すブロック図である。

【図 1 5】

優先処理に対する負荷と、音声認識処理に割り当て可能な CPU パワーとの関係を示す図である。

【図 1 6】

行動決定機構部 52 の処理を説明するためのフローチャートである。

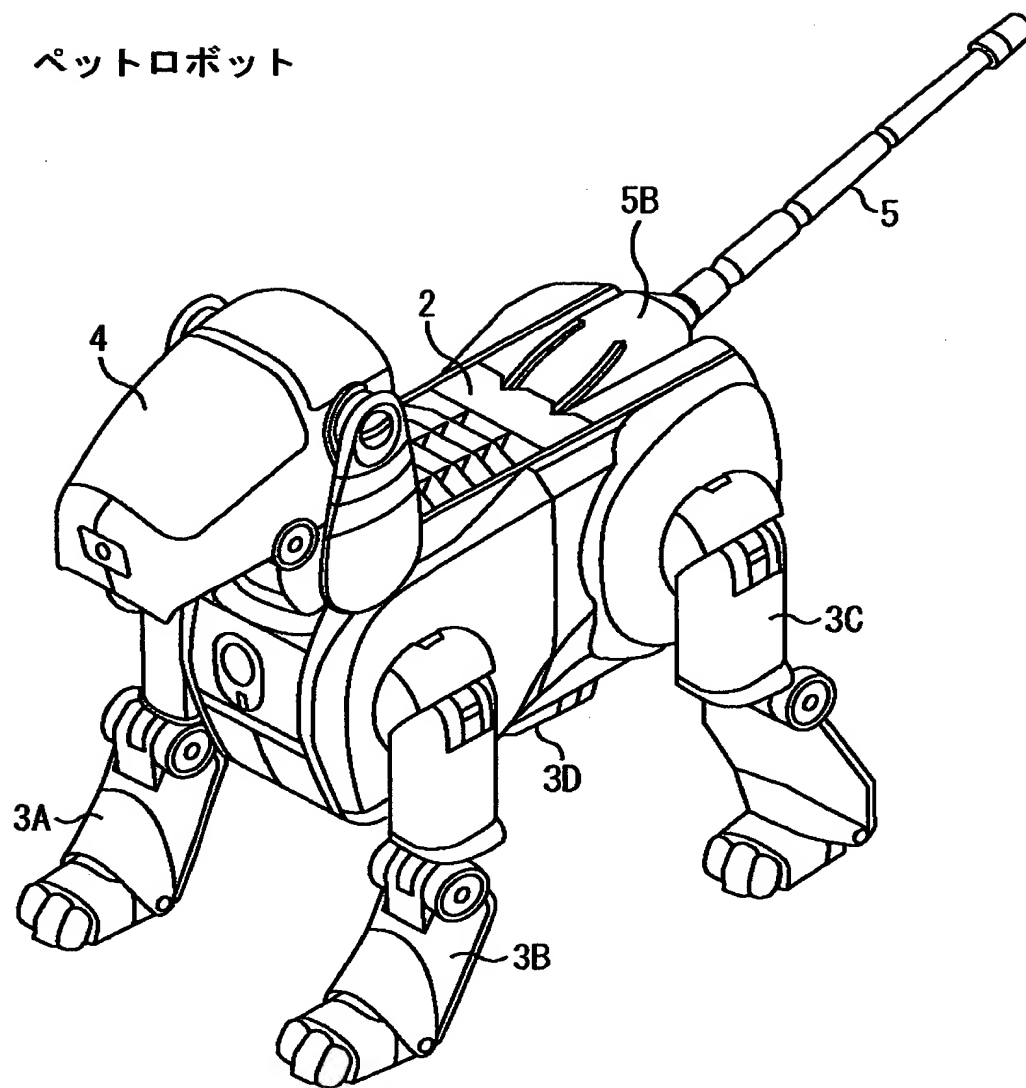
【符号の説明】

10 コントローラ, 10A CPU, 10B メモリ, 15 マイク,  
16 CCDカメラ, 17 タッチセンサ, 18 スピーカ, 21  
AD変換部, 22 特徴抽出部, 23 マッチング部, 24 音響モデル  
記憶部, 25 辞書記憶部, 26 文法記憶部, 31 テキスト生成部,  
32 規則合成部, 33 DA変換部, 34 辞書記憶部, 35 解析  
用文法記憶部, 36 音素片記憶部, 50 センサ入力処理部, 50A  
音声認識部, 50B 画像認識部, 50C 圧力処理部, 51 感情／本  
能モデル部, 52 行動決定機構部, 53 姿勢遷移機構部, 54 制御  
機構部, 55 音声合成部

【書類名】 図面

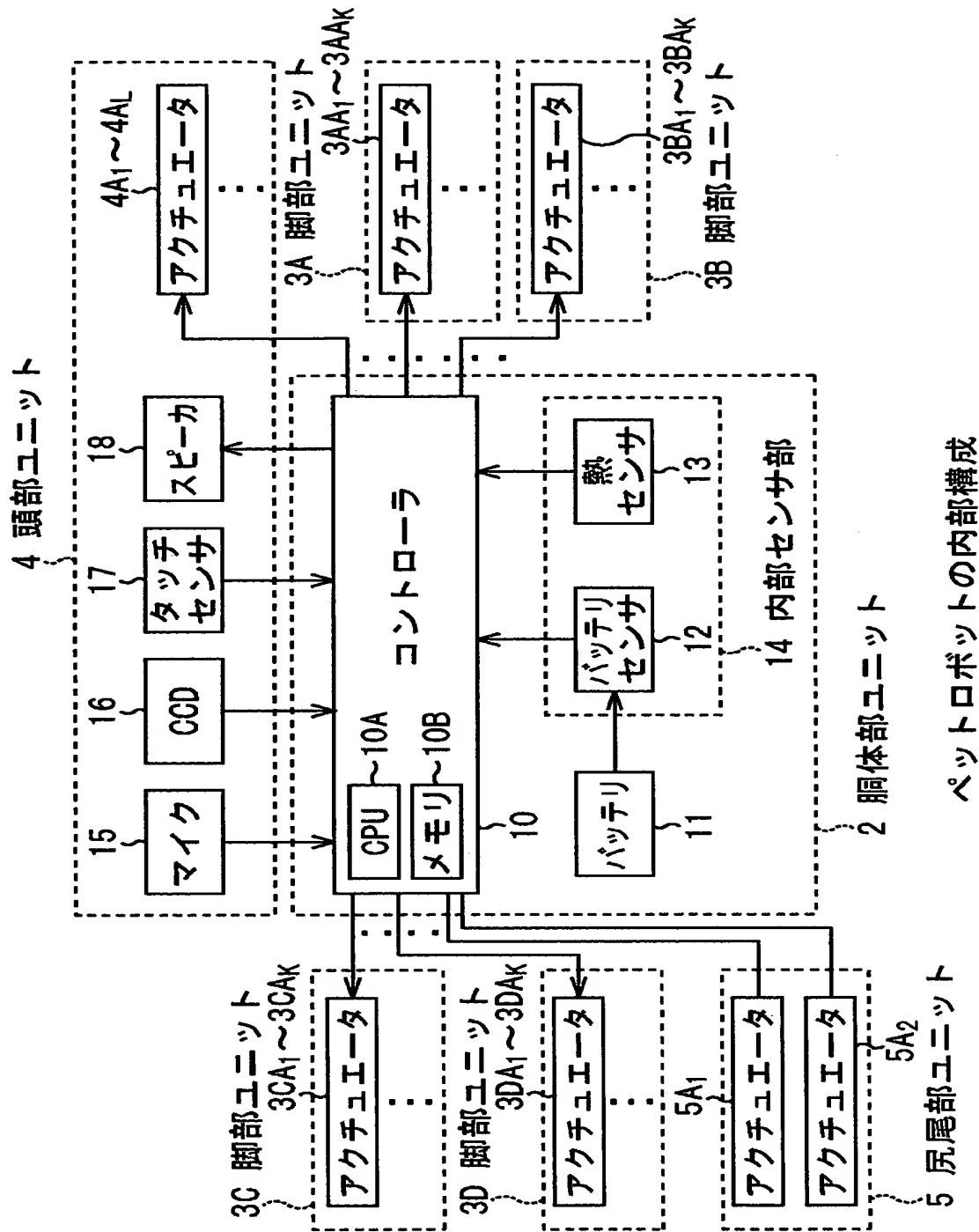
【図 1】

ペットロボット



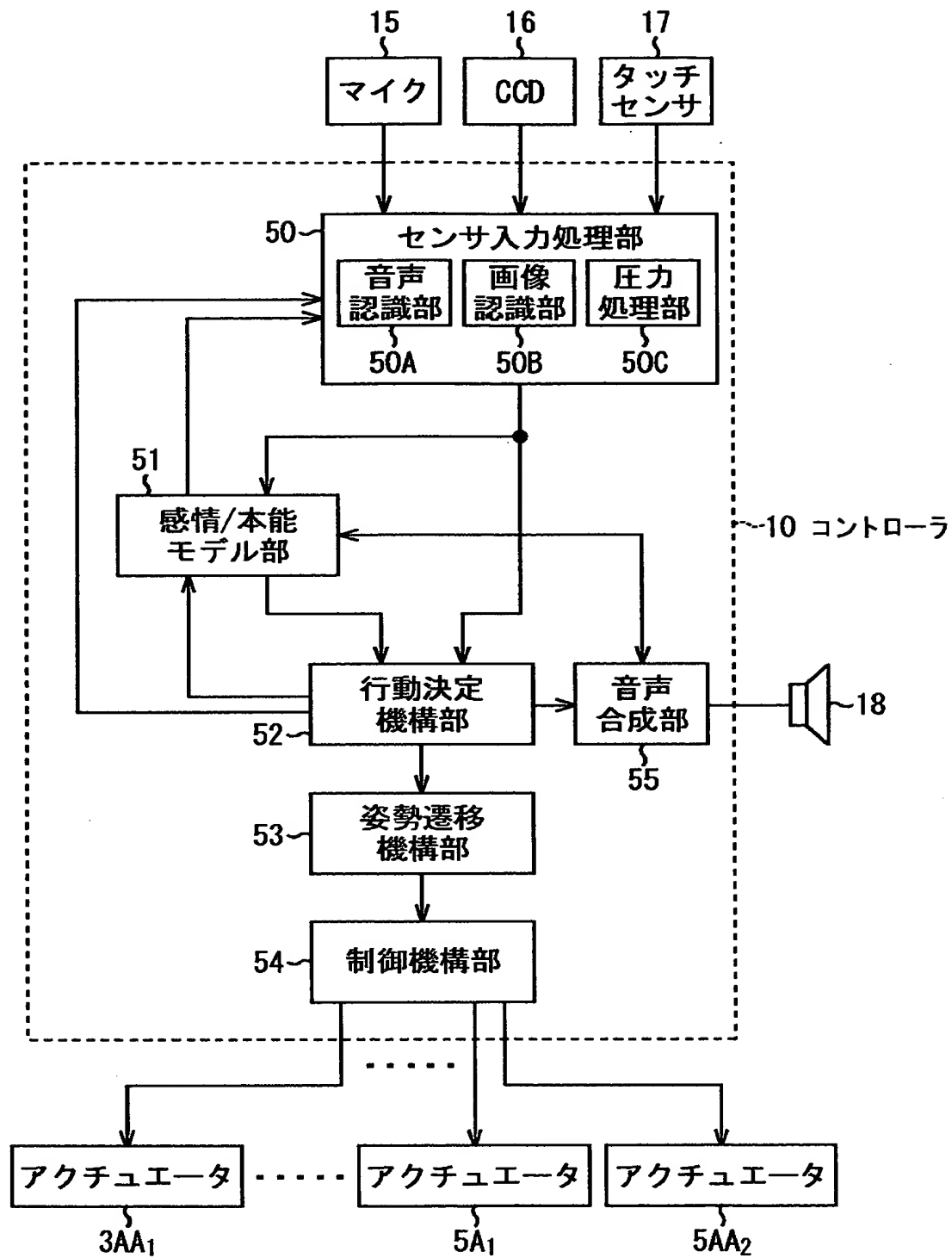
ペットロボットの外観構成

【図 2】

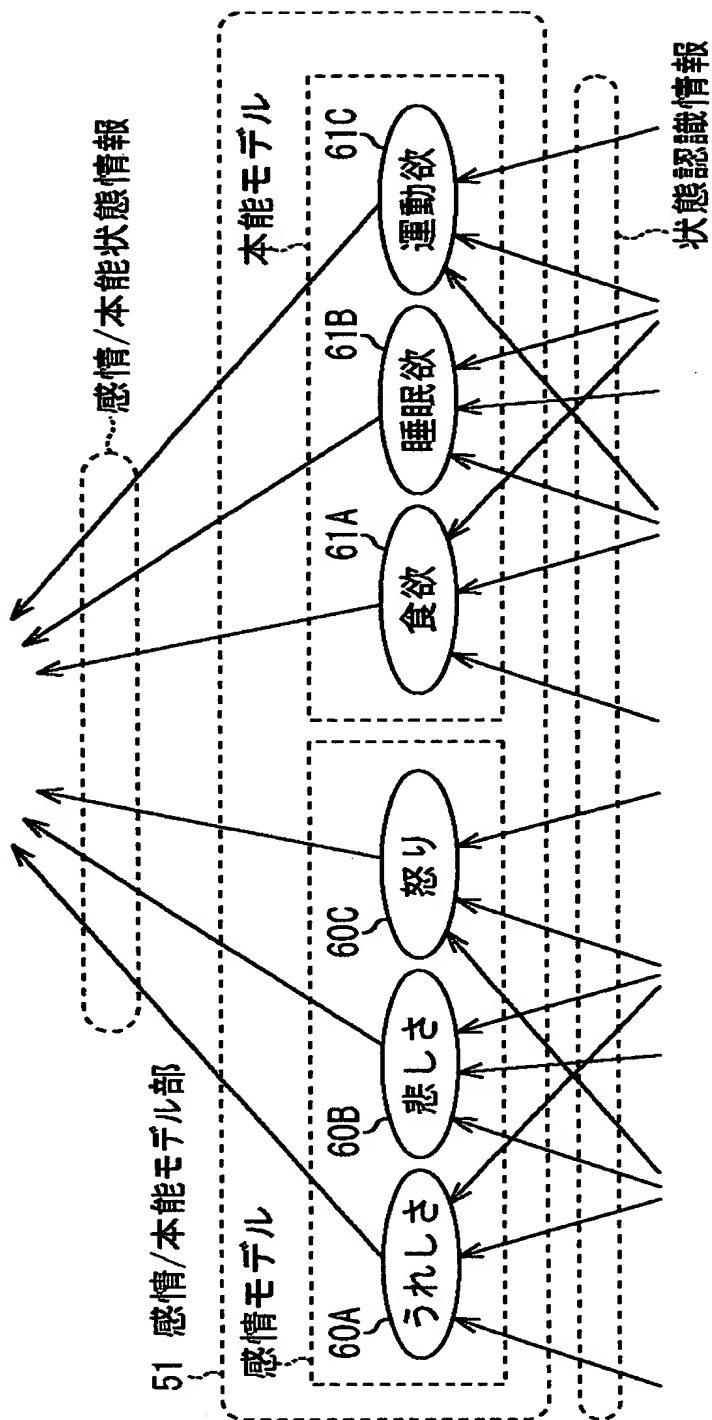


ペットロボットの内部構成

【図 3】

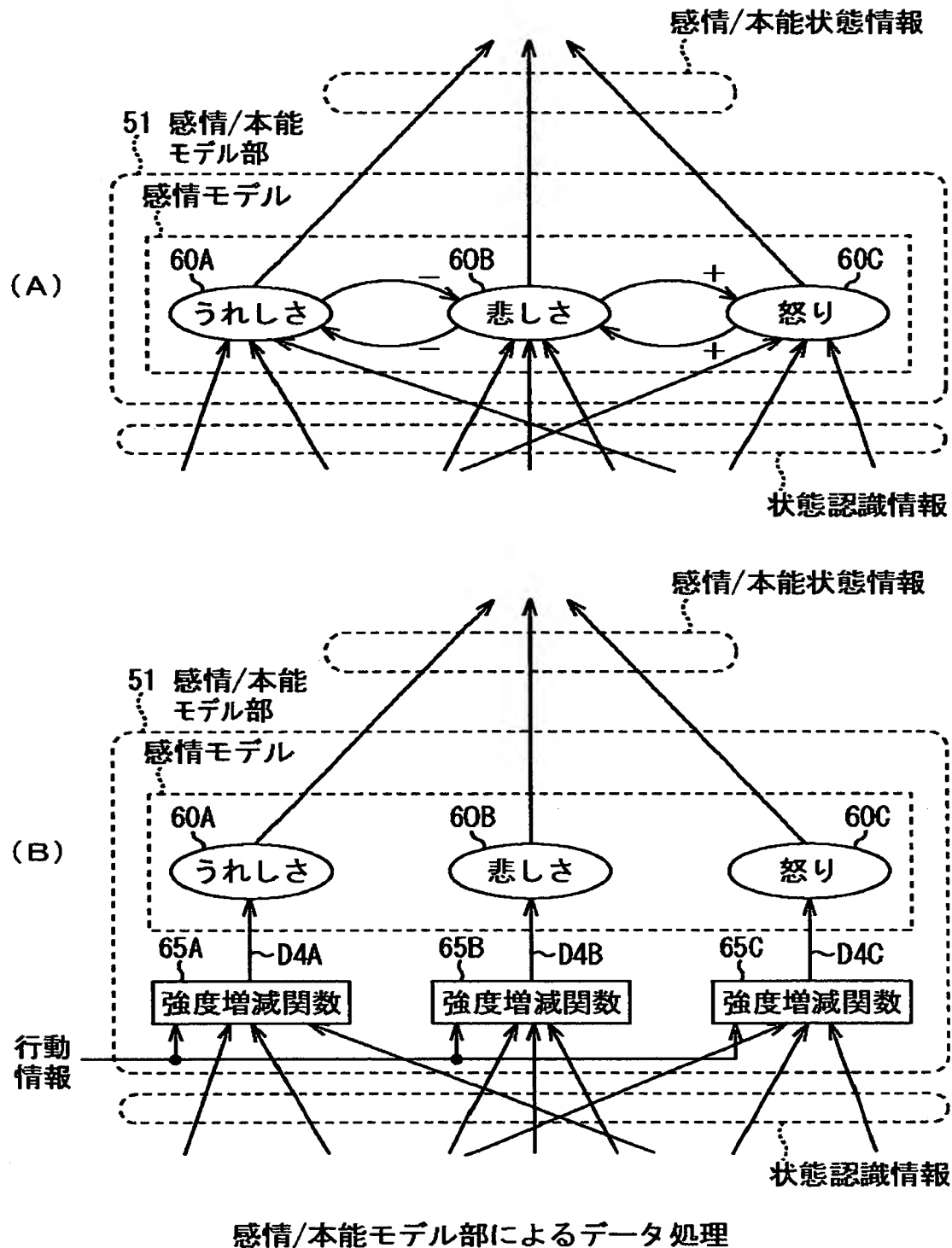


【図 4】

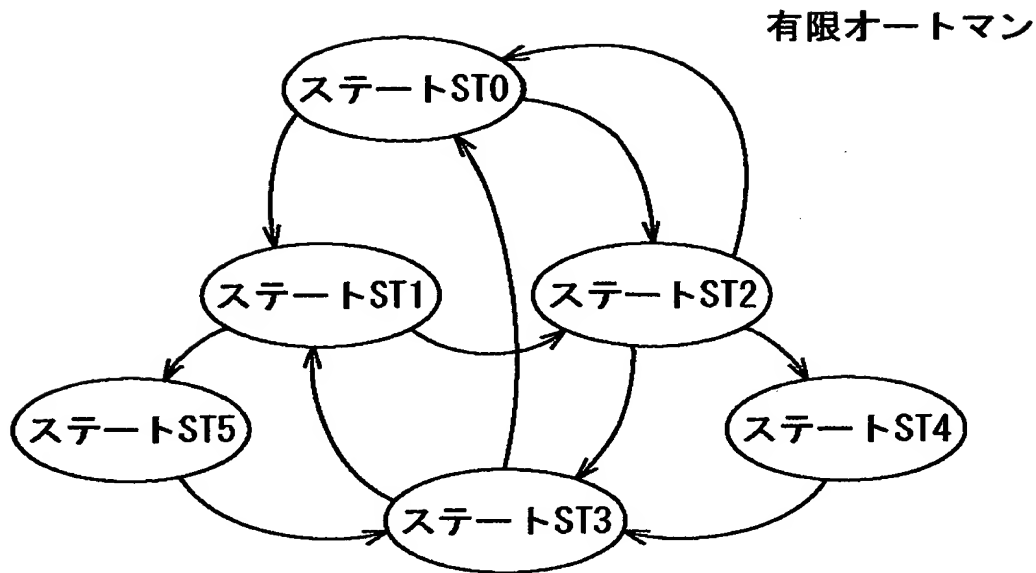


感情/本能モデル部によるデータ処理

【図 5】



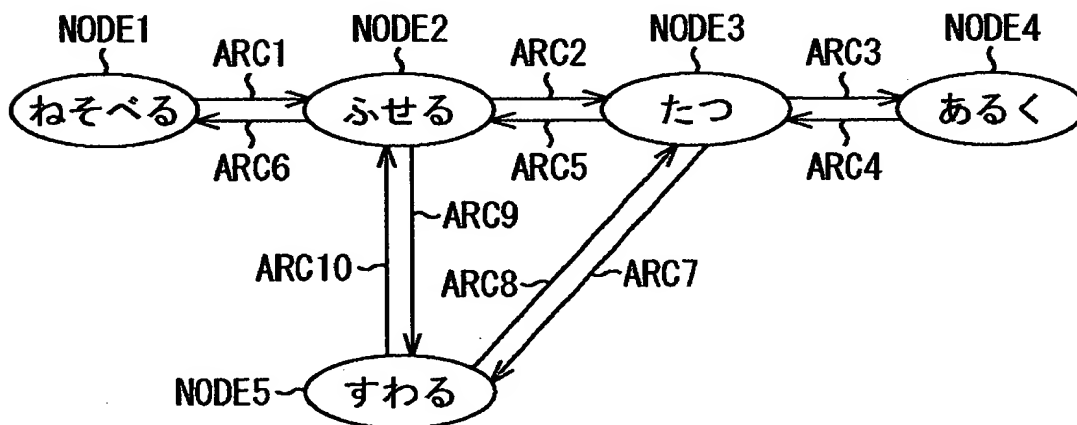
【図 6】



有限オートマンの状態遷移図

【図 7】

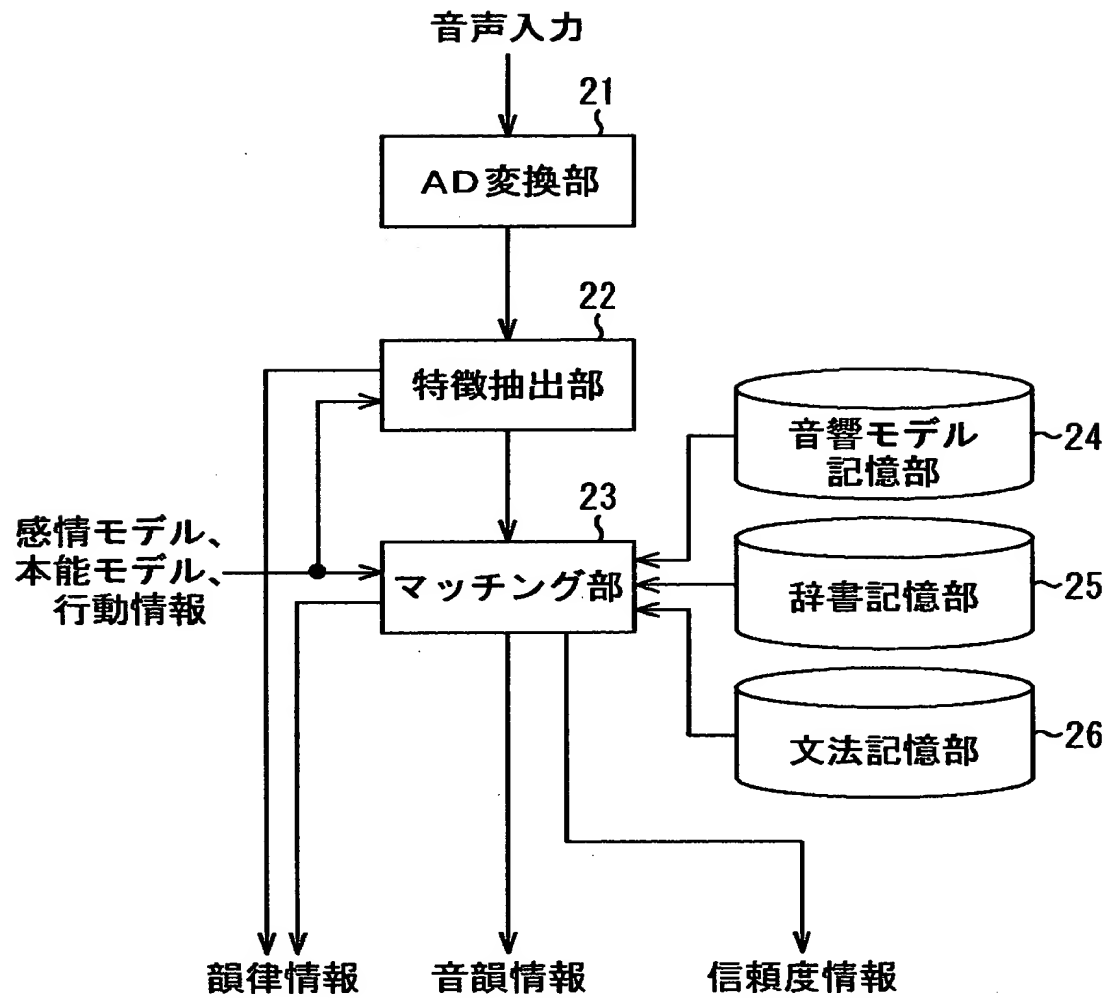
有向グラフ



姿勢遷移図

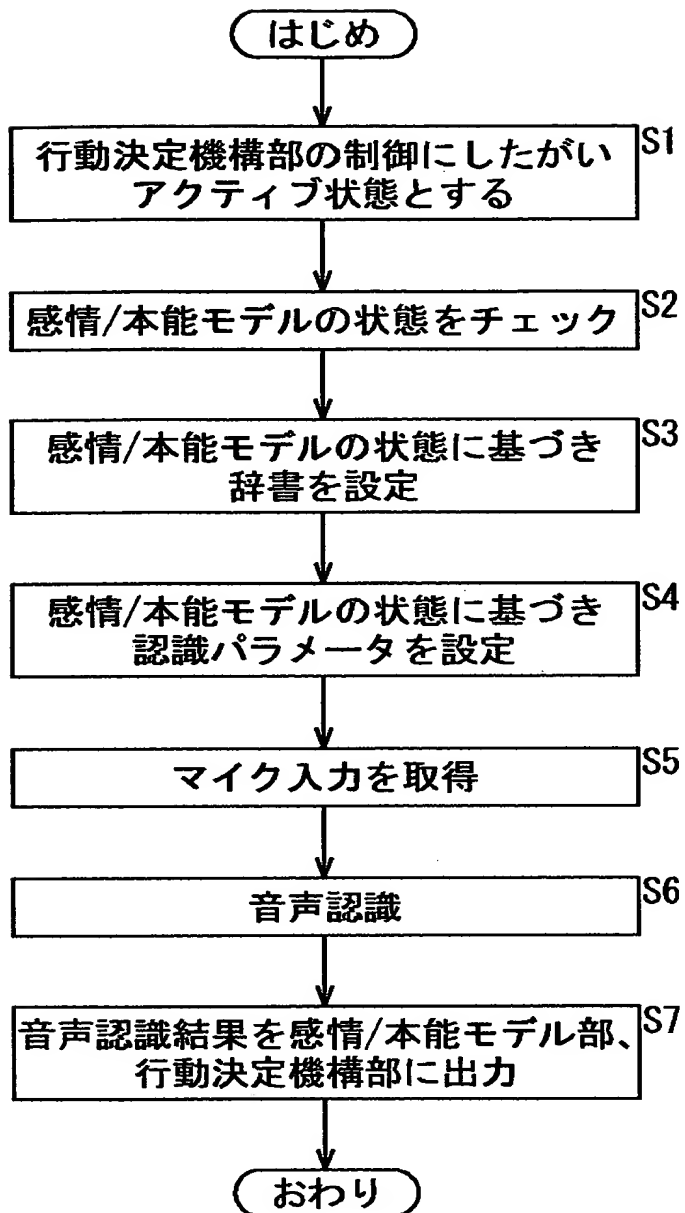


【図 8】

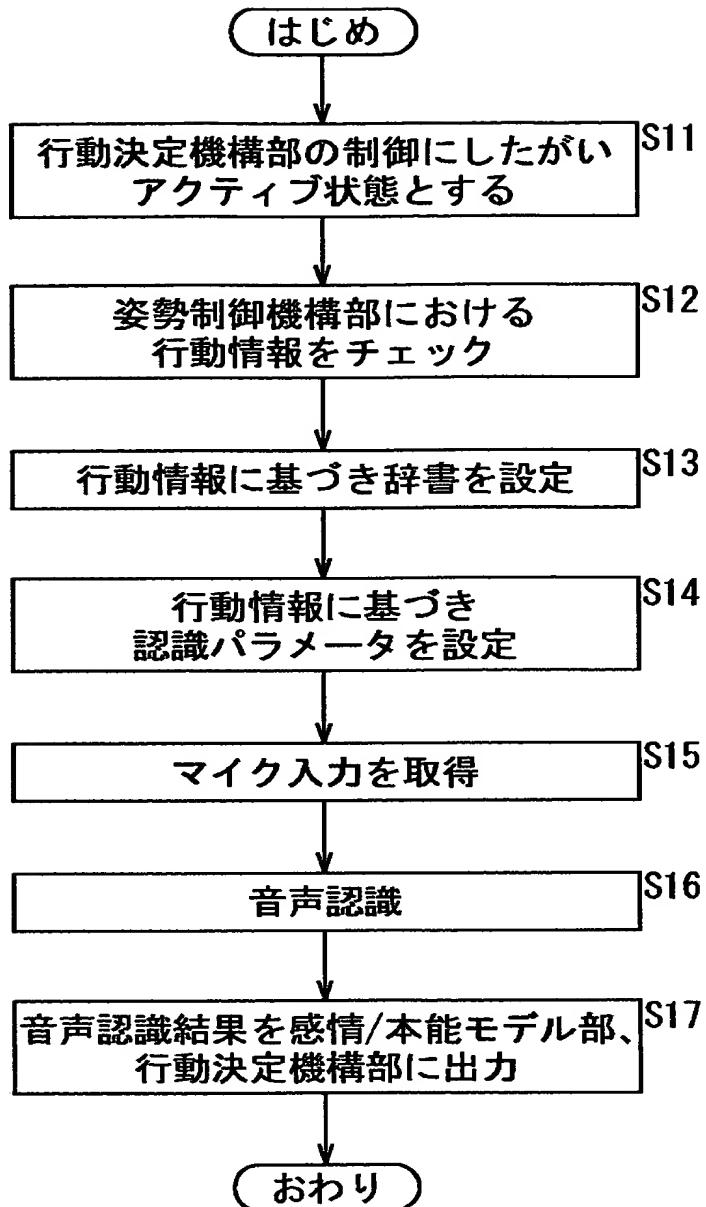


音声認識部 50A

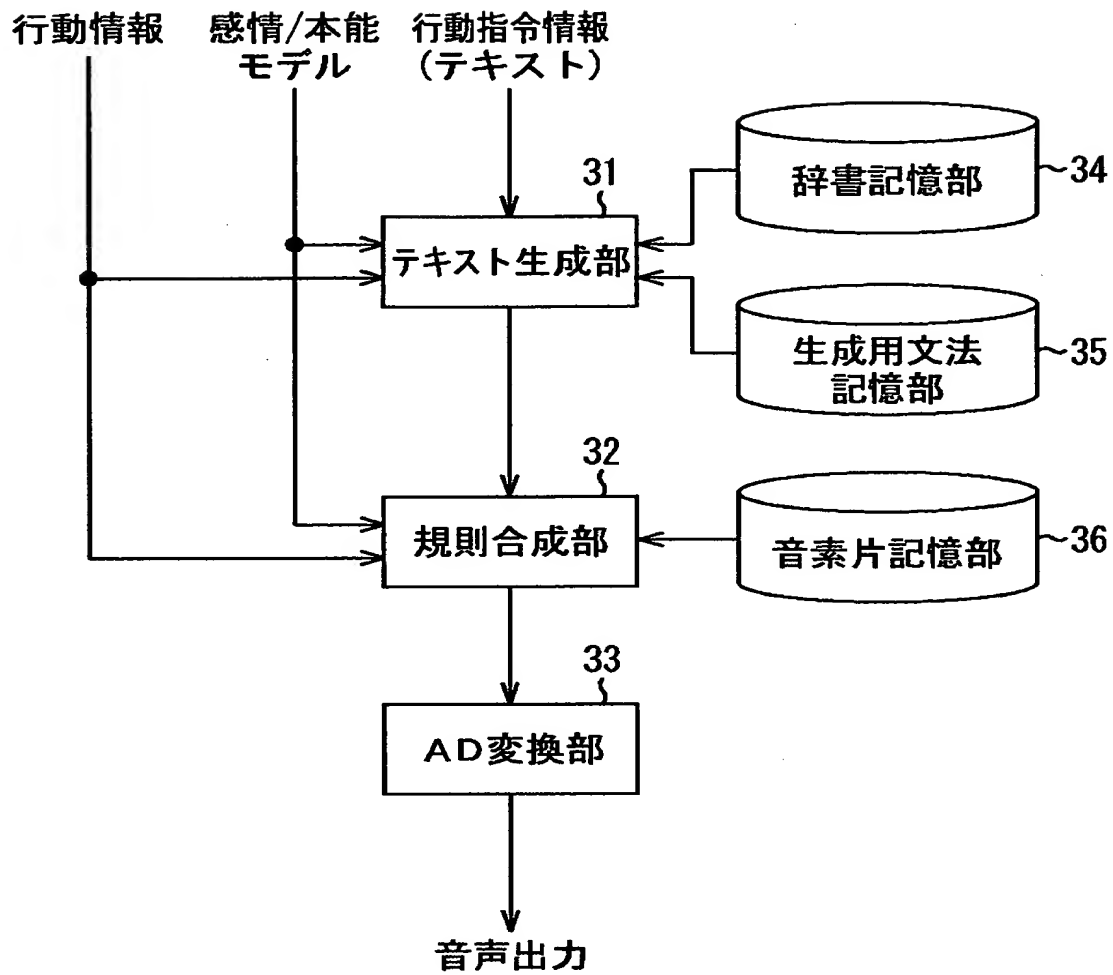
【図 9】



【図 1 0】

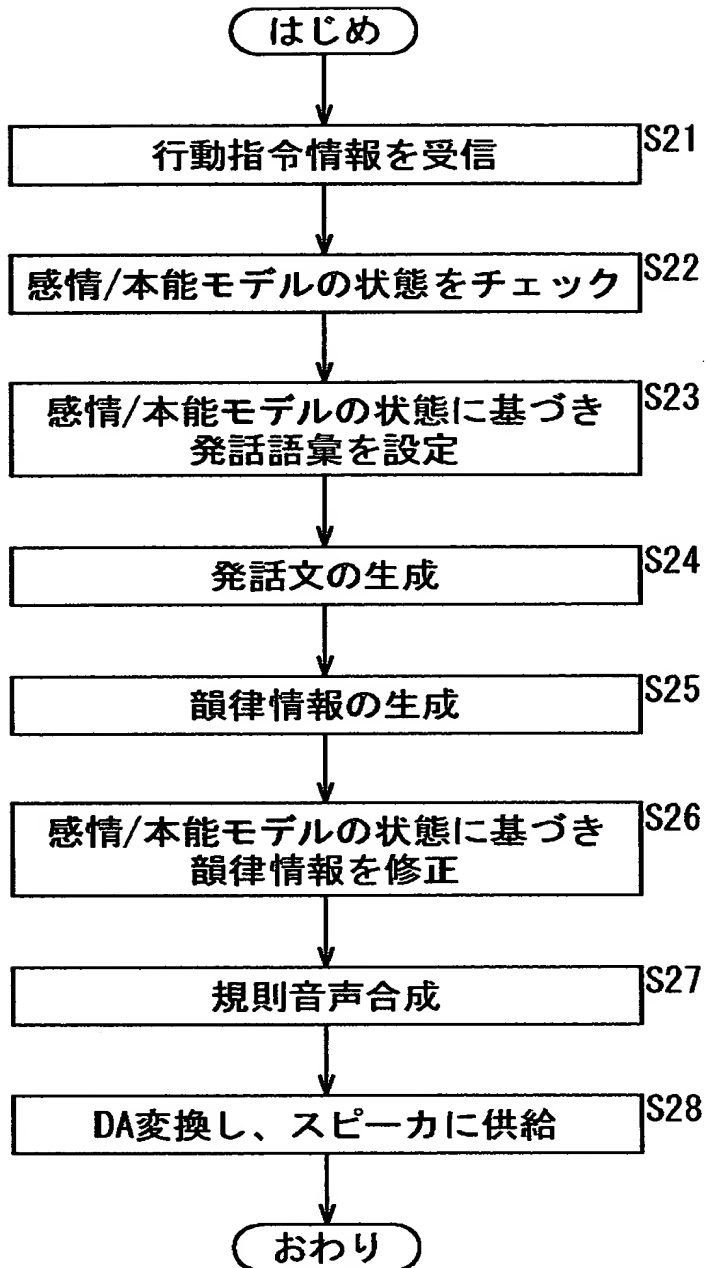


【図 1 1】

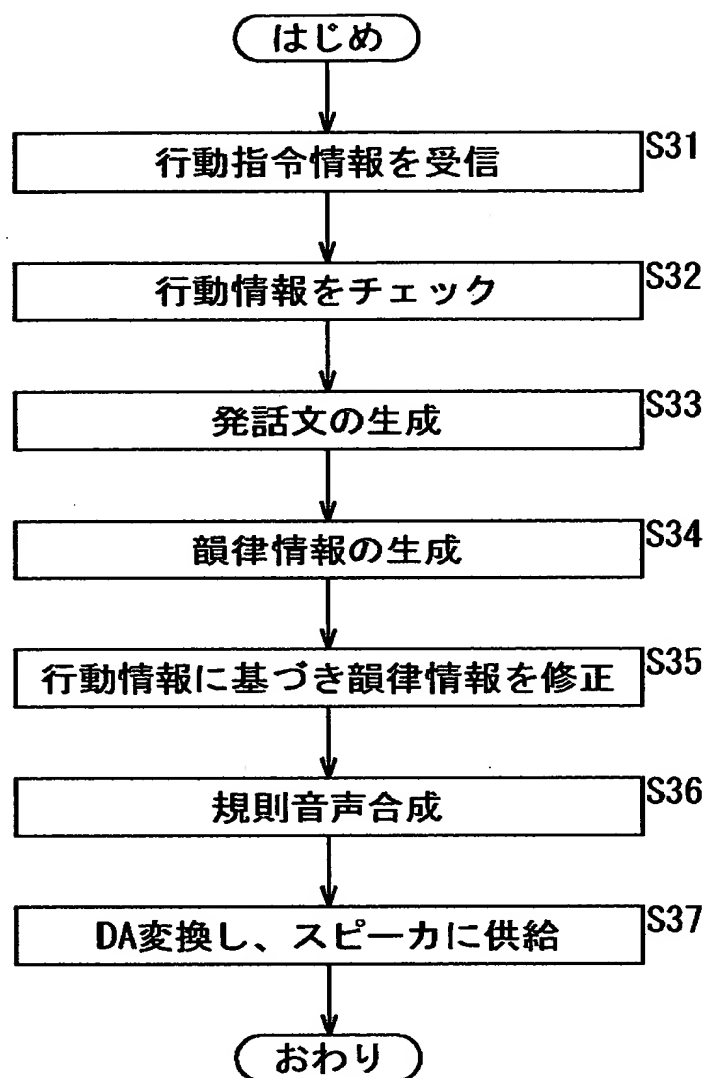


音声合成部 55

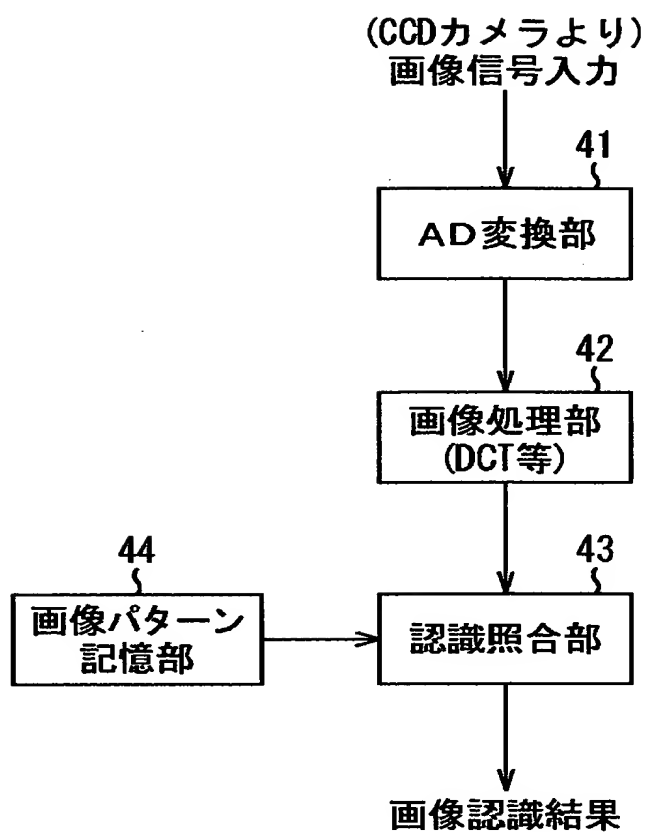
【図 1 2】



【図 1 3】



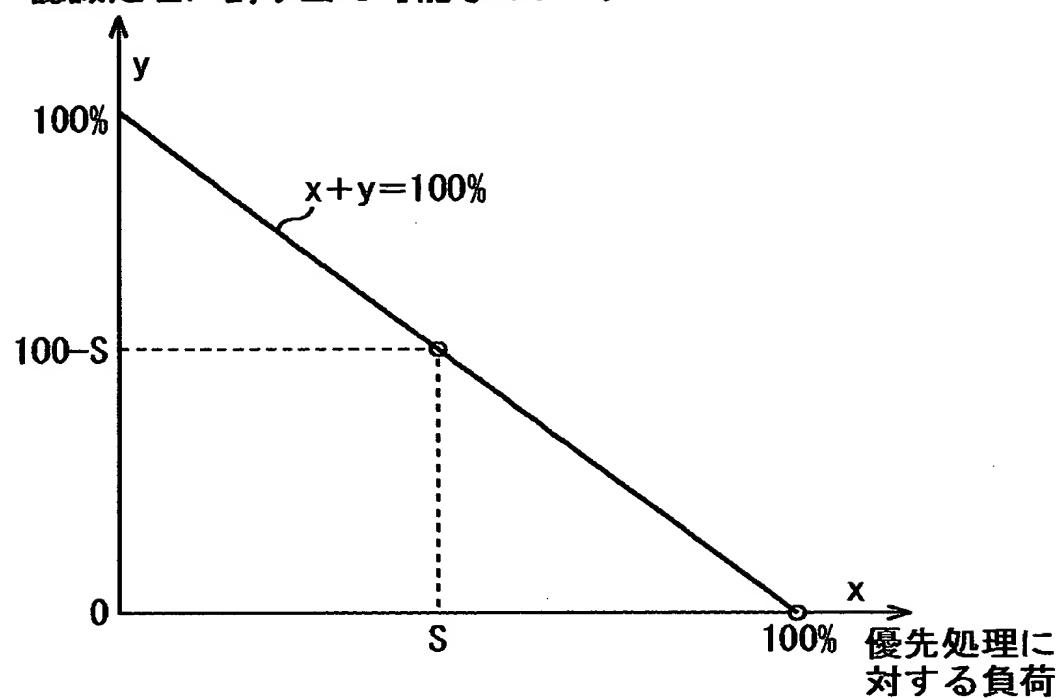
【図 1 4】



画像認識部 50B

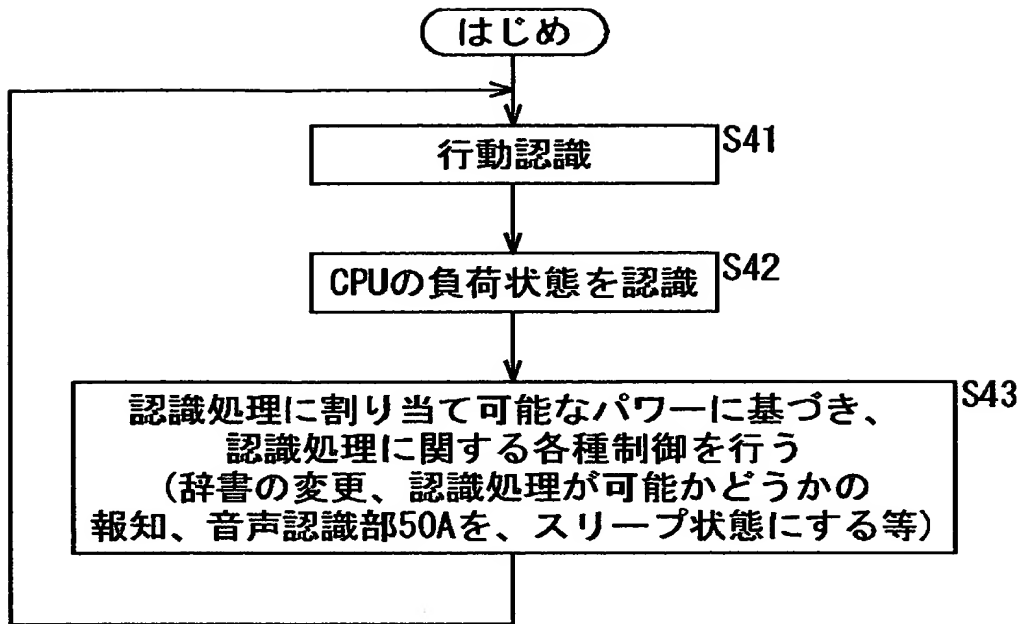
【図 1 5】

認識処理に割り当て可能なCPUパワー





【図 1 6】



【書類名】 要約書

【要約】

【課題】 エンタテインメント性の高いロボットを提供する。

【解決手段】 音声合成部 5 5 では、感情／本能モデル部 5 1 におけるロボットの感情の状態に基づいて、音声合成処理が行われる。即ち、例えば、ロボットの感情の状態が「怒っていない」ことを表しているときには、音声合成部 5 5 では、合成音「何ですか？」が生成される。一方、ロボットの感情の状態が「怒っている」ことを表しているときには、音声合成部 5 5 では、その怒りを表現する「何だよ！」が生成される。

【選択図】 図 3

出 願 人 履 歴 情 報

識別番号 [000002185]

1. 変更年月日	1990年 8月30日
[変更理由]	新規登録
住 所	東京都品川区北品川6丁目7番35号
氏 名	ソニー株式会社